

CCE estimation of factor-augmented regression models with more factors than observables

Citation for published version (APA):

Karabiyik, H., Urbain, J. R. Y. J., & Westerlund, J. (2014). *CCE estimation of factor-augmented regression models with more factors than observables*. Maastricht University, Graduate School of Business and Economics. GSBE Research Memoranda No. 007 <https://doi.org/10.26481/umagsb.2014007>

Document status and date:

Published: 01/01/2014

DOI:

[10.26481/umagsb.2014007](https://doi.org/10.26481/umagsb.2014007)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Hande Karabiyik,
Jean-Pierre Urbain,
Joakim Westerlund

**CCE Estimation of Factor-
Augmented Regression
Models with more Factors
than Observables**

RM/14/007

GSBE

Maastricht University School of Business and Economics
Graduate School of Business and Economics

P.O Box 616
NL- 6200 MD Maastricht
The Netherlands

CCE ESTIMATION OF FACTOR-AUGMENTED REGRESSION MODELS WITH MORE FACTORS THAN OBSERVABLES

Hande Karabiyik
Maastricht University
The Netherlands

Jean-Pierre Urbain*
Maastricht University
The Netherlands

Joakim Westerlund
Deakin University
Australia

February 26, 2014

Abstract

This paper considers estimation of factor-augmented panel data regression models with homogenous slope coefficients. One of the most popular approaches towards this end is the pooled common correlated effects (CCE) estimator of Pesaran (Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* **74**, 967–1012, 2006). For this estimator to be consistent at the usual \sqrt{NT} rate, where N and T denote the number of cross-section and time series observations, respectively, the number of factors cannot be larger than the number of observables. This is a problem in the typical application involving only a small number of regressors. The current paper proposes a simple extension to the CCE procedure by which the requirement can be relaxed. The CCE approach is based on taking the cross-section average of the observables as an estimator of the common factors. The idea put forth in the current paper is to consider not only the average but also other cross-section combinations. The asymptotic properties of the resulting combination-augmented CCE (C³E) estimator are provided and verified in small samples using Monte Carlo simulation.

JEL Classification: C12; C13; C33.

Keywords: Factor-augmented panel regressions; common factor models; principal components; cross-sectional averages; cross-sectional dependence.

*Corresponding author: Maastricht University, School Business and Economics, Department of Quantitative Economics, Tongersestraat 53, 6211 EM, Netherlands. Telephone: +31 43 388 3660. E-mail address: j.urbain@maastrichtuniversity.nl.

1 Introduction

Consider the scalar and $m \times 1$ vector of observable panel data variables $y_{i,t}$ and $x_{i,t}$, where $i = 1, \dots, N$ and $t = 1, \dots, T$ indexes the cross-sectional and time series dimensions, respectively. The data generating process (DGP) of the $T \times 1$ vector $y_i = (y_{i,1}, \dots, y_{i,T})'$ is similar to the DGP of Pesaran (2006), and is given by

$$y_i = x_i \beta + e_i, \quad (1)$$

$$e_i = F \lambda_i + \epsilon_i, \quad (2)$$

where $x_i = (x'_{i,1}, \dots, x'_{i,T})'$ is $T \times m$, β is a $m \times 1$ vector of slope coefficients, $F = (F'_1, \dots, F'_T)'$ is a $T \times r$ matrix of common factors with λ_i being the associated $r \times 1$ vector of factor loadings, and $\epsilon_i = (\epsilon_{i,1}, \dots, \epsilon_{i,T})'$ is a $T \times 1$ vector of idiosyncratic errors. If the model includes unit-specific fixed effects, then y_i , x_i , e_i , F and ϵ_i are simply the correspondingly (time) demeaned variables.

The above model is the prototypical pooled panel regression with a factor error structure, in which ϵ_i is independent of x_i .¹ If F is also independent of x_i , then (1) is nothing but a static panel data regression with exogenous regressors, which can be estimated consistently using least squares (LS). If, however, x_i is correlated with F , then consistency will be lost. To allow for this possibility, we follow Pesaran (2006) and assume that

$$x_i = F \Lambda'_i + \eta_i, \quad (3)$$

where Λ_i is a $m \times r$ loading matrix and $\eta_i = (\eta_{i,1}, \dots, \eta_{i,T})'$ is a $T \times m$ matrix of idiosyncratic errors. By combining (1)–(3),

$$w_i = F C_i + u_i, \quad (4)$$

where $w_i = (y_i, x_i) = (w_{i,1}, \dots, w_{i,T})'$ is $T \times (m+1)$, $w_{i,t} = (y_{i,t}, x'_{i,t})'$ is $(m+1) \times 1$, $C_i = (\Lambda'_i \beta + \lambda_i, \Lambda'_i)$ is $r \times (m+1)$, and $u_i = (u'_{i,1}, \dots, u'_{i,T})' = (\eta_i \beta + \epsilon_i, \eta_i)$ is $T \times (m+1)$. Thus, (1)–(3) can be rewritten equivalently as a static factor model for w_i , which is convenient because it means that the common component of the data can be estimated using existing methods for such models (see Chudik and Pesaran, 2013, for a recent survey).² In this paper, however,

¹One difference between the above DGP and the one considered by Pesaran (2006) is that here β is assumed to be homogenous; see Section 2 for a more detailed discussion.

²In Section 4 of the present paper we present some Monte Carlo result that enable comparison with the principal components-based estimator of Bai (2009), which is arguably the closest competitor of the CCE approach.

we focus on the CCE approach of Pesaran (2006), which has become very popular in the empirical literature with a large number of applications. The approach has also attracted much interest in the econometric literature where it has been shown to work under very general conditions, including models with weak factors, dynamic models and even models with non-stationary data (see, for example, Chudik et al., 2011; Chudik and Pesaran, 2013; Kapetanios et al., 2011; Pesaran et al., 2013).

As is well known from the classical common factor literature, F and C_i are not separately identifiable, suggesting that the best that one can hope for is consistent estimation of the space spanned by F . The idea of Pesaran (2006) is to make use of the cross-section variation to estimate this space. A natural way to accomplish this is to take the cross-section average, giving $\bar{w}_t = \bar{C}'F_t + \bar{u}_t$, where \bar{C} , \bar{w}_t and \bar{u}_t are the cross-section averages of C_i , $w_{i,t}$ and $u_{i,t}$, respectively. Hence, since $\bar{u}_t = o_p(1)$ (as $N \rightarrow \infty$), we have that $\bar{C}'F_t = \bar{w}_t - \bar{u}_t = \bar{w}_t + o_p(1)$. This suggests using \bar{w}_t as an estimator of $\bar{C}'F_t$, a strategy that would seem to require

$$rk(\bar{C}') = r \leq m + 1, \quad (5)$$

where $rk(\bar{C}')$ denotes the rank of \bar{C}' . Hence, the number of observables must be at least as large as the number of factors. The idea behind the CCE approach is then to estimate β as the LS slope on x_i in a regression of y_i onto x_i and \bar{w}_t .

As Pesaran (2006) points out, the condition in (5) can actually be relaxed even within the original CCE framework. However, as we explain in detail in Section 3 of the current paper, this requires imposing additional restrictive conditions on λ_i and Λ_i , which, if false, may well render the CCE estimator inconsistent. Hence, even if (5) can in principle be relaxed, in most situations of practical relevance this is not the case.

The number of regressors, m , is usually a small number that is given by economic theory (and/or previous empirical evidence). By contrast, economic theory is usually not very informative regarding the number of factors, r (see, for example, Eberhardt et al., 2013). Therefore, the theoretically implied value of m has typically little or nothing to do with r . This is important because within the original CCE framework choosing m also means restricting r , and in many applications there is little or no reason to believe that this number should be less than or equal to $m + 1$. However, in practice this aspect is almost always ignored.

In the current paper we take this shortcoming as our starting point. The purpose is to provide a simple modification of the original CCE approach allowing (but not requiring)

$r > m + 1$. The idea is to consider not only the equal-weighted average, but also other combinations of $w_{1,t}, \dots, w_{N,t}$. In particular, by considering k such combinations we can allow for $k(m + 1) \geq m + 1$ common factors. In addition to the larger number of factors that can be allowed, the new approach also enables one to consider separately the selection of m and r . The asymptotic properties of the proposed combination-augmented CCE (C^3E) estimator are studied under the assumption that $N, T \rightarrow \infty$ with $\sqrt{T}/N \rightarrow 0$, which is less restrictive than the $T/N \rightarrow 0$ condition of Pesaran (2006). Some Monte Carlo results are also provided to suggest the asymptotic properties are borne out well in small samples.

The remainder of the paper is organized as follows. Section 2 gives the assumptions, which are used in Section 3 to derive the asymptotic distribution of the proposed C^3E estimator. When $T/N \rightarrow \tau < \infty$ but $\sqrt{T}/N \rightarrow 0$ the estimator is biased. As a response to this, we propose using bias correction, a procedure that is shown to be effective provided that $\sqrt{T}/N \rightarrow 0$ and $\sqrt{N}/T \rightarrow 0$. As a solution to the practical problem of how to pick the appropriate combinations, an information criterion (IC)-based selection rule is proposed. Section 4 focuses on the finite-sample accuracy of the theory provided in Section 3. Section 5 offers a conclusion.

A word on notation. $tr(A)$ and $\|A\| = \sqrt{tr(A'A)}$ denote the trace and Frobenius (Euclidean) norm, respectively, of the matrix A . Also, $M_A = I_T - A(A'A)A'$ for any T -rowed matrix A . $M < \infty$ denotes a generic positive number. Finally, \sim and \rightarrow_p signify asymptotic equivalence and convergence in probability, respectively.

2 Assumptions

The assumptions placed on ϵ_i , η_i and F may be summarized in the following way.

Assumption 1.

- (i) $\epsilon_{i,t} = \phi_i(L)\epsilon_{i,t}$, where $\epsilon_{i,t}$ is independent and identically distributed (iid) with $E(\epsilon_{i,t}) = 0$, $E(\epsilon_{i,t}^2) = 1$, $E(\epsilon_{i,t}^4) \leq M$, $\phi_i(L) = \sum_{j=0}^{\infty} \phi_{ji}L^j$, $\sigma_{\epsilon,i}^2 = E(\epsilon_{i,t}^2) = \sum_{j=0}^{\infty} \phi_{ji}^2 > 0$, and $\sum_{j=0}^{\infty} j|\phi_{ji}| \leq M$;
- (ii) $\eta_{i,t} = \Phi_i(L)v_{i,t}$, where $v_{i,t}$ is iid with $E(v_{i,t}) = 0$, $E(v_{i,t}v_{i,t}') = I_m$, $E(\|v_{i,t}\|^4) \leq M$, $\Phi_i(L) = \sum_{j=0}^{\infty} \Phi_{ji}L^j$, $\Sigma_{\eta,i} = E(\eta_{i,t}\eta_{i,t}') = \sum_{j=0}^{\infty} \Phi_{ji}\Phi_{ji}'$ is positive definite, and $\sum_{j=0}^{\infty} j\|\Phi_{ji}\| \leq M$;

- (iii) F_t is covariance stationary such that $E(||F_t||^4) \leq M$ and $E(F_t F_t') = \Sigma_F$ is positive definite;
- (iv) λ_i and Λ_i are either random such that $E(||\lambda_i||) \leq M$ and $E(||\Lambda_i||) \leq M$, or non-random such that $||\lambda_i|| < \infty$ and $||\Lambda_i|| < \infty$;
- (v) $\varepsilon_{i,t}$, $v_{i,t}$ and F_t are mutually independent.

Remark 1. Assumption 1 is very similar to Assumptions 1–3 in Pesaran (2006), and we therefore refer to this paper for a discussion. The main difference is that while in Pesaran (2006, Assumption 3) λ_i and Λ_i are assumed to be iid with mean λ and Λ , respectively, and positive definite covariance matrices, here λ_i and Λ_i can be either random in a general (non-iid) way or non-random.³ Another difference is that while we only consider the case when β is homogenous, Pesaran (2006) also considers the more general case when β varies randomly across the cross-section. However, as we explain in Section 3, while this random specification of β is certainly more general when it comes to the types of slope behavior that can be permitted, there is also a “cost” to this greater generality.

It is convenient to cast everything in terms of stacked N -dimensional vectors. Let us therefore define the $N(m+1) \times N(m+1)$ matrix $\Sigma_u = \text{diag}(\Sigma_{u,1}, \dots, \Sigma_{u,N})$, where

$$\Sigma_{u,i} = E(u_{i,t} u_{i,t}') = \begin{bmatrix} \beta' \Sigma_{\eta,i} \beta + \sigma_{\varepsilon,i}^2 & \beta' \Sigma_{\eta,i} \\ \Sigma_{\eta,i} \beta & \Sigma_{\eta,i} \end{bmatrix}.$$

For each column in w_i , we consider k cross-section combinations, as given by the $T \times (m+1)k$ matrix $N^{-1} \sum_{i=1}^N w_i Z_i$, where $Z_i = (I_{m+1} \otimes z_i')$ is $(m+1) \times (m+1)k$ and z_i is $k \times 1$. In stacked form, we have $Z = (Z_1', \dots, Z_N')' = (z \otimes I_{m+1})$, a $N(m+1) \times (m+1)k$ matrix, where $z = (z_1, \dots, z_N)'$ is $N \times k$. The columns of z will henceforth be referred to as “combinations” with z being the “combination matrix”. The corresponding $N(m+1) \times r$ matrix stacking C_i is given by $C = (C_1, \dots, C_N)'$. The columns of z can be deterministic and/or stochastic, provided that Assumption 2 is satisfied, where $\bar{H} = N^{-1} Z' C$ is $(m+1)k \times r$.

Assumption 2.

- (i) $rk(\bar{H}) = r \leq (m+1)k$ for all $N < \infty$ and $\bar{H} \rightarrow_p H$ as $N \rightarrow \infty$, where $rk(H) = r$ and $||H|| < \infty$;

³Either way, the type of cross-section dependence permitted is of the strong form. However, in analogy with Chudik et al. (2011, Section 4) some of the factors can also be non-strong without affecting the results.

- (ii) Z_i is either deterministic such that $\|Z_i\| \leq M$, or stochastic such that $E(\|Z_i\|^2) \leq M$;
- (iii) Z is independent of $u_{i,t}$;
- (iv) $N^{-1}Z'\Sigma_u Z \rightarrow_p \Theta$ as $N \rightarrow \infty$, where Θ is a $(m+1)k \times (m+1)k$ positive definite matrix.

Remark 2. In terms of the above notation, the original CCE approach is obtained by setting $k = 1$, $\bar{H} = \bar{C}'$ and $z_i = 1$, and in this case Assumption 2 is the same as in Pesaran (2006). Note in particular how in this case Assumption 2 (i) reduces to $rk(\bar{H}) = rk(\bar{C}') = r \leq m+1$, requiring that the number of observables must be as least as large as the number of factors. By considering $k \geq 1$ combinations we can accommodate $(m+1)k \geq m+1$ factors.

Remark 3. The equal-weighted average is not the only way to combine the data. In fact, as Pesaran (2006) points out, any weighting scheme satisfying granularity condition (14) in his paper will do. However, while recognizing the fact that the weights does not have to be equal, it is still just one combination/weighted average per observable that is being considered. The contribution of the present paper is the consideration of multiple combinations, which is important, because it relaxes the requirement that $m+1 \geq r$.

3 Asymptotic results

3.1 The C³E estimator

As already mentioned, since F and C_i are not separately identifiable, F can only be estimated up to a matrix rotation. The proposed estimator \hat{F} of $F\bar{H}$ is given by

$$\hat{F} = \frac{1}{N} \sum_{i=1}^N w_i Z_i = \frac{1}{N} \sum_{i=1}^N (w_i \otimes z_i'),$$

whose dimension is $T \times (m+1)k$. The resulting estimator of β is given by

$$\hat{\beta} = \left(\sum_{i=1}^N x_i' M_{\hat{F}} x_i \right)^{-1} \sum_{i=1}^N x_i' M_{\hat{F}} y_i.$$

Theorem 1. Under Assumptions 1 and 2, as $N, T \rightarrow \infty$ with $\sqrt{T}/N \rightarrow 0$ and $T/N \rightarrow \tau < \infty$,

$$\sqrt{NT}(\hat{\beta} - \beta) \sim N(0, \bar{\Sigma}_\eta^{-1} \bar{W} \bar{\Sigma}_\eta^{-1}) + \bar{\Sigma}_\eta^{-1} \sqrt{\tau} B,$$

where \sim signifies asymptotic equivalence, and

$$\begin{aligned}\bar{W} &= \frac{1}{N} \sum_{i=1}^N \sigma_{\epsilon,i}^2 \Sigma_{\eta,i}, \\ B &= B_1 - B_2 - B_3, \\ B_1 &= \frac{1}{N} \sum_{i=1}^N \Lambda_i (\bar{H}' \bar{H})^{-1} \bar{H}' \Theta \bar{H} (\bar{H}' \bar{H})^{-1} \lambda_i, \\ B_2 &= \frac{1}{N} \sum_{i=1}^N \Sigma_{\eta,i}(\beta, I_m) Z_i \bar{H} (\bar{H}' \bar{H})^{-1} \lambda_i, \\ B_3 &= \frac{1}{N} \sum_{i=1}^N \sigma_{\epsilon,i}^2 \Lambda_i (\bar{H}' \bar{H})^{-1} \bar{H}' Z_i' (1, 0_m)'. \end{aligned}$$

Remark 4. Note that if $k = 1$ and $z_i = 1$, then $\hat{F} = N^{-1} \sum_{i=1}^N (w_i \otimes z_i') = \bar{w}$, and we are back in the original CCE approach of Pesaran (2006).

Remark 5. The fact that the C³E estimator is biased when $T/N \rightarrow \tau > 0$ is not covered by the theory provided by Pesaran (2006), who assumes $T/N \rightarrow 0$. Theorem 1 can therefore be seen as a generalization of the results of Pesaran (2006, Theorem 4) to the case when $T/N \rightarrow 0$ is violated, making it more relevant for applied work. The presence of bias is, however, not unexpected given the results of Bai (2009, Theorem 3), and Westerlund and Urbain (2013a, Theorem 1).

Remark 6. The C³E estimator considered here is based on “within” pooling, whereby the data are summed over the cross-section before taking the ratio. Another approach is to use “between” pooling, in which case the ratio is taken prior to summing over the cross-section. Pesaran (2006) considers both types of pooling. However, since in his Monte Carlo study within pooling generally leads to the best performing estimator, in this paper we only consider this type.

The remainder of this subsection is devoted to a discussion of the necessity of the rank condition in (5). While under Assumptions 1 and 2 original CCE requires $m + 1 \geq r$, as is made clear in Pesaran (2006), by strengthening the assumptions placed on λ_i it is actually possible to allow $m + 1 < r$ (without the need for additional combinations). To illustrate this, suppose that instead of (1) we have $y_i = x_i \beta_i + e_i$, where $\beta_i = \beta + v_i$ and v_i is iid with mean zero and positive definite covariance matrix Σ_v . Also, v_i is independent of all other random

elements of the DGP. Suppose first that F is known and that $m + 1 = r$, in which case

$$\begin{aligned}
\sqrt{N}(\hat{\beta} - \beta) &= \left(\frac{1}{NT} \sum_{i=1}^N x_i' M_F x_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^N x_i' M_F (x_i v_i + F \lambda_i + \epsilon_i) \\
&= \left(\frac{1}{NT} \sum_{i=1}^N \eta_i' M_F \eta_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' M_F \eta_i v_i + o_p(1) \\
&= \bar{\Sigma}_\eta^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \Sigma_{\eta,i} v_i + o_p(1) \\
&\sim \left(\bar{\Sigma}_\eta^{-1} \frac{1}{N} \sum_{i=1}^N \Sigma_{\eta,i} \Sigma_v \Sigma_{\eta,i} \bar{\Sigma}_\eta^{-1} \right)^{1/2} N(0, I_m),
\end{aligned}$$

where the last result follows from applying a central limit law to $N^{-1/2} \sum_{i=1}^N \Sigma_{\eta,i} v_i$. Thus, in this case normality is a direct consequence of the assumed randomness of β_i . A similar result applies in case $m + 1 < r$, but then normality requires that λ_i and Λ_i have similar random coefficient representations as β_i . In fact, as pointed out by Westerlund and Urbain (2013b), in this case one also has to assume that Λ_i and λ_i are mutually independent, which seems like a rather restrictive assumption. For example, when regressing investments on savings, as is commonly done in the literature on the so-called “Feldstein–Horioka puzzle”, a common shock that increases savings is going to push interest rates down and investments up, suggesting that Λ_i and λ_i should be negatively correlated. Thus, while the requirement that $m + 1 \geq r$ can be relaxed also within the original CCE framework, this does not come free of charge. Then there is also the dependence of the asymptotic distribution of $\sqrt{N}(\hat{\beta} - \beta)$ on Σ_v , which makes for relatively complicated covariance estimation. Finally, if β_i is allowed to vary, the rate of convergence is reduced, from \sqrt{NT} to \sqrt{N} , which is true also in case of a violation of $m + 1 \geq r$.

One may think that the above mentioned reduction in the rate of consistency is due to the fact that β_i is allowed to vary, and that imposing $\beta_1 = \dots = \beta_N = \beta$ would prevent this from happening, regardless of whether $m + 1 \geq r$ or $m + 1 < r$.⁴ However, this is not the case. The reason for this is easily appreciated by replacing F in the above expansion of $\sqrt{N}(\hat{\beta} - \beta)$ with $\hat{F} = \bar{w}$. By imposing $v_1 = \dots = v_N = 0$ and noting that $(NT)^{-1/2} \sum_{i=1}^N x_i' M_{\bar{w}} \epsilon_i = O_p(1)$, the numerator $\sqrt{NT}(\hat{\beta} - \beta)$ reduces to

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^N x_i' M_{\bar{w}} (x_i v_i + F \lambda_i + \epsilon_i) = \frac{1}{\sqrt{NT}} \sum_{i=1}^N x_i' M_{\bar{w}} F \lambda_i + O_p(1).$$

⁴It is not clear from Pesaran (2006) whether one can have $\beta_1 = \dots = \beta_N = \beta$, while at the same time permitting $m + 1 < r$.

On the one hand, if $m + 1 \geq r$, then, since in such cases $\hat{F} = F\bar{C} + o_p(1)$, $x'_i M_{\bar{w}} F = x'_i M_{F\bar{C}} F = 0$, and so we obtain $\sqrt{NT}(\hat{\beta} - \beta) = O_p(1)$. Hence, provided that $m + 1 \geq r$, imposing $\beta_1 = \dots = \beta_N = \beta$ restores \sqrt{NT} -consistency. On the other hand, if $m + 1 < r$, then $x'_i M_{\bar{w}} F = O_p(T)$, and therefore

$$\sqrt{NT}(\hat{\beta} - \beta) = \sqrt{T} \left(\frac{1}{NT} \sum_{i=1}^N x'_i M_{\bar{w}} x_i \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N T^{-1} x'_i M_{\bar{w}} F \lambda_i + O_p(1),$$

whose order is determined by the order of the first term on the right, which in turn depends on λ_i and Λ_i . If λ_i is iid independently of Λ_i , then the first term is $O_p(\sqrt{T})$, whereas if λ_i is non-iid and/or correlated with Λ_i , then the same term is $O_p(\sqrt{NT})$. Thus, the rate of consistency is \sqrt{N} , at best, and if λ_i is non-iid and/or correlated with Λ_i , then $\hat{\beta}$ is even inconsistent. The proposed estimator is not only very simple, but is also \sqrt{NT} -consistent regardless of the specification of λ_i and Λ_i , provided that Assumptions 1 and 2 are satisfied.

3.2 Bias correction

As pointed out by Bai (2009), an obvious solution to the problem with bias in the case when $T/N \rightarrow \tau > 0$ is to use bias correction. Let us therefore define the following bias-adjusted version of $\hat{\beta}$:

$$\hat{\beta}_{BA} = \hat{\beta} - N^{-1} \hat{\Sigma}_\eta^{-1} \hat{B}.$$

Here \hat{B} is B with β , $\Sigma_{\eta,i}$ and $\sigma_{\epsilon,i}^2$ replaced by $\hat{\beta}$, $\hat{\Sigma}_{\eta,i}$ and $\hat{\sigma}_{\epsilon,i}^2$, respectively, where $\hat{\Sigma}_{\eta,i}$ and $\hat{\sigma}_{\epsilon,i}^2$ are the usual heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimators based on $M_{\hat{F}} x_i$ and $(y_i - x_i \hat{\beta} - \hat{F} \hat{\lambda}_i)$, respectively, and $\hat{\Sigma}_\eta = N^{-1} \sum_{i=1}^N \hat{\Sigma}_{\eta,i}$. The estimators $\hat{\lambda}_i$ and $\hat{\Lambda}_i$ of λ_i and Λ_i , respectively, are obtained by simply picking the appropriate elements in \hat{C}_i , which can be obtained from a LS regression of $w_{i,t}$ onto \hat{F}_t .

Proposition 1. *Under Assumptions 1 and 2, as $N, T \rightarrow \infty$ with $\sqrt{T}/N \rightarrow 0$ and $\sqrt{N}/T \rightarrow 0$,*

$$\sqrt{NT}(\hat{\beta}_{BA} - \beta) \sim N(0, \bar{\Sigma}_\eta^{-1} \bar{W} \bar{\Sigma}_\eta^{-1}).$$

According to Proposition 1, asymptotically the bias correction is successful. Moreover, the correction does not contribute to the limiting variance, which is the same as in Theorem 1.

3.3 Selecting the combinations

A problem in applications is how to pick the combinations, z . The most obvious approach is to exploit if there are natural candidates. z is ideally chosen to be uncorrelated with $\epsilon_{i,t}$, but highly correlated with C , measuring the exposure of each individual unit to the common shocks. As such, natural candidates for z are actually not difficult to find. For example, if $w_{i,t}$ includes GDP and/or inflation, then z might be a vector of trade shares, or, if $w_{i,t}$ includes stock prices, then z might be the book-to-market ratio and/or the earnings-price ratio (aggregated over the whole time period). In fact, z can even include preliminary consistent estimates of (the space spanned by) C . The only requirement is that the rate of consistency must be at least \sqrt{N} , which is sufficiently relaxed to enable estimation by, for example, principal components (see Bai, 2003, Theorem 2). Deterministic instruments are even simpler to come by. Again, if $w_{i,t}$ includes GDP, then z might be a vector of distance rankings, or, if $w_{i,t}$ includes stock prices, then z might be industrial classification. Also, as Chudik et al. (2011) show, the cross-sectional average can be quite effective in mopping up cross-section dependence. Thus, a vector of ones is always a good starting point when constructing z .

Even if candidates for z are easy to find, in practice there will most likely be uncertainty regarding which candidates to use. If this is the case, then z can be estimated by an IC of the form

$$IC(s) = V(s, \hat{F}^s) + sg(N), \quad (6)$$

where \hat{F}^s is \hat{F} based on s combinations, $V(s, \hat{F}^s) = (NT)^{-1} \sum_{i=1}^N (y_i' M_{x_i} M_{\hat{F}^s} M_{x_i} y_i)$ is the sum of squared residuals, and $g(N)$ is a penalty function. To fix ideas, suppose for sake of argument that m is known and that the columns of z are ordered according to relevance, as measured by the (cross-section) correlation with (the unobserved) columns of C . The problem of determining z therefore reduces to the problem of consistent estimation of k , the dimension of z . The IC estimator \hat{k} of k is given by

$$\hat{k} = \arg \min_{s=0, \dots, k_{max}} IC(s),$$

where $k_{max} \geq k$. By Assumption 2, $k \geq \lfloor r/(m+1) \rfloor$, where $\lfloor x \rfloor$ is the largest integer less than x . Hence, we do not propose to use \hat{k} as an estimator of the number of factors, r , but as an estimator of the minimal number of columns of z required to approximate the underlying factor structure.

Proposition 2. Under Assumptions 1 and 2, if $g(N) \rightarrow 0$ and $Ng(N) \rightarrow \infty$, as $N \rightarrow \infty$,

$$P(\hat{k} = k) \rightarrow 1.$$

Remark 7. Bai and Ng (2002) proposes several IC that are appropriate in the context of principal components estimation approximate common factor models. The class of IC considered here is very similar. The only difference is the penalty, which is here assumed not to depend on T . The reason for this difference is that while the rate of consistency of the principal components estimator of F depends on both N and T , the rate of consistency of \hat{F} only depends on N .

Remark 8. In order to appreciate the implications of Proposition 2 for the asymptotic distribution of $\sqrt{NT}(\hat{\beta} - \beta)$ it is useful to treat $\hat{\beta}$ as a function k . Let us therefore write $\hat{\beta}^k$ for $\hat{\beta}$. Clearly,

$$\begin{aligned} P[\sqrt{NT}(\hat{\beta}^{\hat{k}} - \beta) \leq \delta] &= P[\sqrt{NT}(\hat{\beta}^{\hat{k}} - \beta) \leq \delta | \hat{k} = k]P(\hat{k} = k) \\ &+ P[\sqrt{NT}(\hat{\beta}^{\hat{k}} - \beta) \leq \delta | \hat{k} \neq k]P(\hat{k} \neq k), \end{aligned}$$

where $\delta > 0$. Because $P(\hat{k} = k) \rightarrow 1$ and $P(\hat{k} \neq k) \rightarrow 0$, the second term on the right-hand side converges to zero, and $P[\sqrt{NT}(\hat{\beta}^{\hat{k}} - \beta) \leq \delta] = 1 + o(1)$. Moreover, conditional on $\hat{k} = k$, $\hat{\beta}^{\hat{k}} = \hat{\beta}^k$. Thus,

$$|P[\sqrt{NT}(\hat{\beta}^{\hat{k}} - \beta) \leq \delta] - P[\sqrt{NT}(\hat{\beta}^k - \beta) \leq \delta]| \rightarrow 0,$$

suggesting that Theorem 1 is unaffected by the estimation of k .

The minimization can be done via grid search over all possible combinations of columns of z ; however, given that the large number of potential candidates, a computationally less demanding procedure seems preferable. To implement the minimization, the following two-step approach may be used:

1. We begin by ordering the candidates. This can be done by considering the cross-section correlation between each element of z_i and each element of C_i . But since the latter is not observed, we propose to using $\bar{w}_i = T^{-1} \sum_{t=1}^T w_{i,t} = C_i' \bar{F} + \bar{u}_i = C_i' \bar{F} + o_p(1)$ to estimate the space spanned by C_i . This practice leads to a $(m+1) \times k$ matrix of correlations, and hence $(m+1)$ potentially different orderings. As a way of summarizing these we propose to take the average of the correlations for each candidate over the elements of \bar{w}_i .

2. Once an ordering of the candidates is obtained, the problem of selecting which combinations to use reduces to the relatively simple problem of selecting the number of combinations, which can be done by using an IC. The specific IC used in this paper can be seen as a version of the IC_{p1} criterion of Bai and Ng (2002) and is given by

$$IC(s) = \ln(V(s, \hat{F}^s)) + sN^{-1} \ln(N).$$

Proposition 2 implies that \hat{k} based on this criterion is consistent for k .

4 Monte Carlo results

In this section we evaluate briefly the small sample properties of the C³E estimator. The DGP used for this purpose can be seen as a restricted version of (1)–(3), and sets $m = 1$, $\beta = -2$ and $(F'_{it}, \eta_{i,t}, \epsilon_{i,t}) \sim N(0, I_{r+2})$.⁵ The difference between the experiments considered lies in how we generate λ_i , Λ_i and z_i . Seven experiments, denoted E1–E7, are considered. In E1–E3, the rank condition in (5) is satisfied, whereas in E4–E7, the condition is violated. In E4 and E5, λ_i and Λ_i are iid and independent of each other, as required in original CCE, whereas in E6 and E7, λ_i and Λ_i are non-iid. Exactly how λ_i , Λ_i and z_i are generated is described in Table A. For each experiment, 19 (N, T) pairs are considered. In the first 16, $N, T = 30, 50, 100, 200$, whereas in last three, $N = \lfloor T^{4/3} \rfloor$. The motivation behind the last three pairs is to assess the performance when the conditions of Theorem 1 and Proposition 1 (that is, $\sqrt{N}/T \rightarrow 0$ and $\sqrt{T}/N \rightarrow 0$) are met. We report the bias and the size of a nominal 5% level t -test. The reported results are based on 5,000 replications and the parameters of the models are kept constant across replications.

The performance of C³E is compared with that of the naive LS estimator where the cross-sectional dependence is ignored, original CCE, and the principal components (PC) estimator of Bai (2009). Four versions of C³E is considered; (i) C³E based on the combinations reported in Table A, (ii) C³E|1 based on the Table A-combinations plus the cross-sectional average of the observables, (iii) C³ES, which is C³E based on IC selection of the Table A-combinations and, (iv) C³ES|1, which is C³E based on IC selection among the combinations reported in Table A plus the cross-sectional averages. The reason for the special treatment of the cross-section averages is that we would like to be able to study separately a violation of the CCE

⁵We also conducted some experiments while allowing for unit-specific fixed effect. As expected, this change did not affect the conclusions. The results are not reported but can be made available upon request.

rank condition in (5) and the one in Assumption 2. As explained in Section 3.3 $C^3E|1$ and $C^3ES|1$ are obtained by simply including in z a vector of ones. These estimators therefore nest original CCE. The selection of the combinations (in C^3ES and $C^3ES|1$) is done as explained in Section 3.3. We begin by ordering the combinations according to their cross-section correlation with the time averages of the observables. Once an ordering is obtained, starting with the full set of combinations, the proposed IC is used to select the appropriate combinations to use.

The results for each experiment are reported in Tables E1–E7. The conclusions that can be drawn from these tables are summarized as follows.

- E1. The aim of this experiment is to evaluate the accuracy of the theoretically predicted bias (as given in Theorem 1) and also to evaluate the performance of the bias-adjusted C^3E estimator.⁶ The results reported in Table E1 suggest that if $N < T$, then the bias of the C^3E estimator is slightly smaller than expected, whereas if $N \geq T$, then the empirical bias is generally very close to the one predicted by theory. As expected, bias correction leads to a considerable improvement for all estimators. The overall best performance is obtained by using bias-adjusted C^3E , followed by CCE and PC.
- E2. The aim of this experiment is to compare the performance of original CCE and C^3E when all the conditions required for both methods to be valid, including (5), are met. All the combinations are highly correlated with the loadings. Under these conditions both CCE and (the versions of) C^3E should perform equally well, which is also reflected in Table E2. CCE tend to perform slightly better, but the difference is not very large, and it gets smaller as N and T increases. As expected given Proposition 2, in large samples C^3ES and $C^3ES|1$ do just as well as C^3E and $C^3E|1$, respectively.
- E3. In this experiment the combinations are generated independently of the loadings, such that Assumption 2 is violated. The expectation is that this should cause C^3E to break down. However, (5) is still satisfied, and therefore CCE is still expected to perform well. Interestingly, this is not what we see in Table E3. Specifically, while the performance of the C^3E estimators is generally worse than in Table E2, the other estimators also perform worse than in E2. In fact, in most of the cases C^3E outperforms the compe-

⁶The theoretical bias of the CCE and C^3E estimators are obtained from Theorem 1. In case of the PC estimator the theoretical prediction is taken from Bai (2009).

tition, including CCE. A comparison of C^3E and $C^3E|1$ shows that when the loadings are generated independently of the combinations, including a vector of ones to the combination matrix decreases the bias for most of the N and T combinations. We also experimented with a subset of invalid combinations and obtained very similar results⁷.

- E4. This experiment is conducted to compare the performance of CCE and the versions of C^3E when (5) is not satisfied, as in this case

$$E(C_i) = \begin{bmatrix} 1.5 & 0.5 \\ 0 & 0 \end{bmatrix}$$

(see Table A). But the factor loadings are still iid. The theory suggests that in this case the rate of consistency of CCE should be reduced from \sqrt{NT} to \sqrt{N} . The results reported in Table E4 are quite suggestive of this; while decreasing in N , the bias is roughly constant in T . We also see that the size of the bias of the CCE is substantially higher when compared to E3. But while the bias is constant, the size distortions of the CCE are actually increasing in T . Moreover, although the bias is generally decreasing in N , the size distortions remain or even increase with increases in N . This is unexpected, because inference based on the CCE t -statistic should be asymptotically valid regardless of the relative expansion rate of N and T (see Pesaran, 2006, Theorem 3). By contrast, the C^3E t -statistic maintains the nominal level well and the accuracy increases with increases in N and/or T . The \sqrt{NT} -consistency of the C^3E estimator is also clearly visible; increasing N and/or T leads a reduction in bias. Furthermore including a vector of ones to the combination matrix (so that C^3E nests original CCE) leads to a decrease in the bias.

- E5. In this experiment (5) is again violated; however, this time the violation is due to the presence of too many factors; $m + 1 = 2 < r = 3$. Also, the loadings are independent of the combinations. The results reported in Table E5 show that while the conclusions regarding the CCE bias are roughly the same as in E4, the size distortions of this estimator are now markedly smaller. However, there is still a tendency for the distortions to increase with T . As expected, since there are now even more invalid combinations than in E3, the C^3E results reported in Table E5 are worse than those reported in Table E3. However, in most of the cases, performance is still better than for CCE.

⁷These unreported results can be made available upon request.

E6. In Table E6,

$$E(C_i) = \begin{bmatrix} 1.7 & 0.2 \\ 0 & 0 \end{bmatrix}$$

and therefore (5) is violated. Also, λ_{2i} and Λ_{1i} are correlated with each other, so the factor loadings are non-iid. As expected, CCE breaks down completely. However, since the combinations are still correlated with the loadings, C^3E still performs well.

E7. All loadings are mean zero and non-iid in this experiment. The CCE estimator is now even more biased than in E7. The C^3E estimator still performs well, as does C^3ES .

5 Conclusion

This paper considers the problem of consistent estimation of a factor-augmented panel regression model in which the number of factors, r , is potentially larger than the number of observables, $m + 1$. The estimator that we propose can be viewed as an extension of the CCE estimator of Pesaran (2006), which is based on using the cross-section averages of the observables as proxies for the latent factors. While CCE does allow $r > m + 1$, it does so at a cost. In particular, it is required that the factor loadings are iid, which in most cases of practical relevance is likely to be violated. Also, even if the assumption is met, the rate of consistency is only \sqrt{N} , as opposed to the usual \sqrt{NT} -rate. In this paper we take this feature of CCE as our starting point. The purpose is to provide a simple extension that preserves \sqrt{NT} -consistency without for that matter requiring iid loadings.

The idea behind the proposed C^3E estimator is to use not only the cross-section average but also other (cross-section) combinations of the observables. By taking $k \geq 1$ such combinations we can allow $k(m + 1) \geq m + 1$ common factors. The new estimator is shown to be \sqrt{NT} -consistent and asymptotically normal under the condition that $T/N \rightarrow \tau < \infty$. This condition is more genal than the $T/N \rightarrow 0$ condition of Pesaran (2006), whose relaxation is shown to have important consequences. In particular, it is shown that the estimator is biased whenever $\tau > 0$. As a response to this, a bias-adjusted C^3E estimator is proposed, which is shown to support asymptotically normal inference under $\sqrt{T}/N \rightarrow 0$ and $\sqrt{N}/T \rightarrow 0$. This is true if the combinations are known. If there is uncertainty over which combinations to use an IC can be used to select the appropriate combinations.

The small-sample performance of the C^3E estimator is examined through a series of

Monte Carlo experiments. The results suggest that whenever the assumptions of Pesaran (2006) are satisfied, the performance of the CCE and C^3E estimators are comparable. If, however, the assumptions are not met, then the C^3E estimator continues to work well while the CCE estimator breaks down. We also find that the proposed bias-adjustment and IC-based combination selection procedures seem to work well, leading to estimators with good small-sample properties.

References

- Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* **71**, 135–173.
- Bai, J. (2009). Panel Data Models with Interactive Fixed Effects. *Econometrica* **77**, 1229–1279.
- Bai, J., and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica* **70**, 191–221
- Pesaran, M. H., and A. Chudik (2013). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. Cambridge Working Papers in Economics 1317, Faculty of Economics, University of Cambridge.
- Chudik, A., M. H. Pesaran and E. Tosetti (2011). Weak and strong cross section dependence and estimation of large panels. *Econometrics Journal* **14**, C45–C90.
- Eberhardt, M., C. Helmers and H. Strauss (2013). Do spillovers matter when estimating private returns to R&D? *The Review of Economics and Statistics* **95**, 436–448.
- Kapetanios, G., M. H. Pesaran and T. Yamagata (2011). Panels with non-stationary multifactor error structures. *Journal of Econometrics* **160**, 326–348.
- Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* **74**, 967–1012.
- Pesaran, M. H., V. Smith and T. Yamagata (2013). Panel unit root tests in the presence of a multifactor error structure. *Journal of Econometrics* **175**, 94–115.
- Westerlund, J, and J.-P. Urbain (2013a). Cross-sectional averages versus principal components. Unpublished manuscript.
- Westerlund, J, and J.-P. Urbain (2013b). On the estimation and inference in factor-augmented panel regressions with correlated loadings. *Economics Letters* **119**, 247–250.

Appendix: Proofs

We start with some notation. The model for $w_{i,t} = (y_{i,t}, x'_{i,t})'$ can be written in matrix notation as

$$w_i = FC_i + u_i, \quad (\text{A1})$$

where $w_i = (y_i, x_i) = (w_{i,1}, \dots, w_{i,T})'$ is $T \times (m+1)$, $F = (F_1, \dots, F_T)'$ is $T \times r$, $C_i = (\Lambda'_i \beta + \lambda_i, \Lambda'_i)$ is $r \times (m+1)$ and $u_i = (u_{i,1}, \dots, u_{i,T})' = (\eta_i \beta + \epsilon_i, \eta_i)$ is $T \times (m+1)$. Alternatively, the model for $w_{i,t}$ can be written as the following N -dimensional system:

$$w_t = CF_t + u_t, \quad (\text{A2})$$

where $w_t = (w'_{1,t}, \dots, w'_{N,t})'$ and $u_t = (u'_{1,t}, \dots, u'_{N,t})'$ are $N(m+1) \times 1$, and $C = (C_1, \dots, C_N)'$ is $N(m+1) \times r$. The matrix notation

$$w = FC' + u \quad (\text{A3})$$

will also be used, where $w = (w_1, \dots, w_N)$ and $u = (u_1, \dots, u_N)$ are $T \times N(m+1)$. In what follows the representations in (A1)–(A3) will be used interchangeably.

Many of the results can be expressed in terms of $(\hat{F} - F\bar{H}')$. It is therefore going to be convenient to introduce some special notation to simplify such expressions. It is therefore useful to define

$$d = \hat{F} - F\bar{H}' = \frac{1}{N} \sum_{i=1}^N u_i Z_i, \quad (\text{A4})$$

whose dimension is given by $T \times (m+1)k$. It is further convenient to write $d = (d_1, \dots, d_T)'$, where

$$d_t = \hat{F}_t - \bar{H}F_t = \frac{1}{N} \sum_{i=1}^N Z'_i u_{i,t} \quad (\text{A5})$$

is $(m+1)k \times 1$.

Before we come to the proof of Theorem 1 we state some useful lemmas.

Lemma A.1. *Under Assumptions 1 and 2,*

$$\frac{1}{T} \sum_{t=1}^T ||d_t||^2 = O_p(N^{-1}).$$

Proof of Lemma A.1.

The proof of Lemma 1 is a simple consequence of the fact that $\|N^{-1/2} \sum_{i=1}^N Z'_i u_{i,t}\| = O_p(1)$, as seen by writing

$$\frac{N}{T} \sum_{t=1}^T \|d_t\|^2 \leq \frac{1}{T} \sum_{t=1}^T \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N Z'_i u_{i,t} \right\|^2 = O_p(1).$$

■

Lemma A.2. *Under the conditions of Lemma A.1,*

$$\|\sqrt{NT}^{-1/2} F' d\| = O_p(1).$$

Proof of Lemma A.2.

The proof is completed by noting that,

$$\sqrt{NT}^{-1/2} F' d = \frac{\sqrt{N}}{\sqrt{T}} \sum_{t=1}^T F_t d'_t = \frac{1}{\sqrt{T}} \sum_{t=1}^T F_t \frac{1}{\sqrt{N}} \sum_{i=1}^N u'_{i,t} Z_i = O_p(1). \quad (\text{A6})$$

■

Lemma A.3. *Under the conditions of Lemma A.1,*

$$NT^{-1} d' d = \Theta + O_p(T^{-1/2})$$

where $\Theta = \lim_{N \rightarrow \infty} N^{-1} Z' \Sigma_u Z$.

Proof of Lemma A.3.

A direct calculation reveals that

$$\begin{aligned} NT^{-1} d' d &= \frac{N}{T} \sum_{t=1}^T d_t d'_t = \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^N Z'_i u_{i,t} u'_{j,t} Z_j \\ &= \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N Z'_i u_{i,t} u'_{i,t} Z_i + \frac{1}{\sqrt{T}} \frac{1}{N\sqrt{T}} \sum_{t=1}^T \sum_{i=1}^N \sum_{j \neq i}^N Z'_i u_{i,t} u'_{j,t} Z_j \\ &= \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N Z'_i u_{i,t} u'_{i,t} Z_i + O_p(T^{-1/2}) \\ &= \Theta + \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N (Z'_i u_{i,t} u'_{i,t} Z_i - \Theta) + O_p(T^{-1/2}) \\ &= \Theta + O_p(T^{-1/2}), \end{aligned} \quad (\text{A7})$$

as was to be shown.

■

Lemma A.4. *Under the conditions of Lemma A.1,*

$$\frac{1}{T} \sum_{i=1}^N \eta'_i d = \frac{1}{N} \sum_{i=1}^N \Sigma_{\eta,i}(\beta, I_m) Z_i + O_p(T^{-1/2}).$$

Proof of Lemma A.4.

By using the fact that $E(\eta_{i,t} u'_{i,t}) = E[\eta_{i,t}(\epsilon_{i,t} + \eta'_{i,t} \beta, \eta'_{i,t})] = (\Sigma_{\eta,i} \beta, \Sigma_{\eta,i}) = \Sigma_{\eta,i}(\beta, I_m)$, we obtain

$$\begin{aligned} \frac{1}{T} \sum_{i=1}^N \eta'_i d &= \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T \eta_{i,t} d'_t = \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T \eta_{i,t} \frac{1}{N} \sum_{j=1}^N u'_{j,t} Z_j = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{j=1}^N \eta_{i,t} u'_{j,t} Z_j \\ &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \eta_{i,t} u'_{i,t} Z_i + \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{j \neq i}^N \eta_{i,t} u'_{j,t} Z_j \\ &= \frac{1}{N} \sum_{i=1}^N \Sigma_{\eta,i}(\beta, I_m) Z_i + \frac{1}{\sqrt{NT}} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T [\eta_{i,t} u'_{i,t} Z_i - \Sigma_{\eta,i}(\beta, I_m) Z_i] \\ &\quad + \frac{1}{\sqrt{T}} \frac{1}{N\sqrt{T}} \sum_{i=1}^N \sum_{t=1}^T \sum_{j \neq i}^N \eta_{i,t} u'_{j,t} Z_j \\ &= \frac{1}{N} \sum_{i=1}^N \Sigma_{\eta,i}(\beta, I_m) Z_i + O_p(T^{-1/2}), \end{aligned} \tag{A8}$$

as required for the proof. ■

Lemma A.5. *Under the conditions of Lemma A.1,*

$$NT^{-1} \epsilon'_i d = \sigma_{\epsilon,i}^2 (1, 0_m) Z_i + O_p(T^{-1/2}).$$

Proof of Lemma A.5.

The proof is implied by $E(\epsilon_{i,t} u'_{i,t}) = E[\epsilon_{i,t}(\epsilon_{i,t} + \eta'_{i,t} \beta, \eta'_{i,t})] = (\sigma_{\epsilon,i}^2, 0)$ and

$$\begin{aligned} &E \left[\left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \sum_{j \neq i}^N \epsilon_{i,t} u'_{j,t} Z_j \right)' \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \sum_{j \neq i}^N \epsilon_{i,t} u'_{j,t} Z_j \right) \right] \\ &= \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \sum_{k \neq i}^N \sum_{j \neq i}^N E(Z'_j u_{j,t} \epsilon_{i,t} \epsilon_{i,s} u'_{k,s} Z_k) \\ &= \frac{1}{T} \sum_{t=1}^T \sum_{k \neq i}^N \sum_{j \neq i}^N E(Z'_j u_{j,t} \epsilon_{i,t} \epsilon_{i,t} u'_{k,t} Z_k) + \frac{1}{T} \sum_{t \neq s}^T \sum_{s=1}^T \sum_{k \neq i}^N \sum_{j \neq i}^N E(Z'_j u_{j,t} \epsilon_{i,t} \epsilon_{i,s} u'_{k,s} Z_k) \\ &= \frac{1}{T} \sum_{t=1}^T \sigma_{\epsilon,i}^2 E(Z'_j u_{j,t} u'_{j,t} Z_j) + \frac{1}{T} \sum_{t=1}^T \sum_{k \neq i}^N \sum_{j \neq i, j \neq k}^N \sigma_{\epsilon,i}^2 E(Z'_j u_{j,t} u'_{k,t} Z_k) = \sigma_{\epsilon,i}^2 \Theta_j, \end{aligned}$$

where $\Theta_j = E(Z_j' \Sigma_{u,j} Z_j)$. Since, $\|\Theta_j\| < \infty$, we have $\|T^{-1/2} \sum_{t=1}^T \sum_{j \neq i}^N \epsilon_{i,t} u'_{j,t} Z_j\| = O_p(1)$, as is clear from

$$\begin{aligned} NT^{-1} \epsilon'_i d &= \frac{1}{T} \sum_{t=1}^T \epsilon_{i,t} \frac{1}{N} \sum_{j=1}^N u'_{i,t} Z_j = \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^N \epsilon_{i,t} u'_{j,t} Z_j = \frac{1}{T} \sum_{t=1}^T \epsilon_{i,t} u'_{i,t} Z_i + \frac{1}{T} \sum_{t=1}^T \sum_{j \neq i}^N \epsilon_{i,t} u'_{j,t} Z_j \\ &= \sigma_{\epsilon,i}^2(1,0) Z_i + \frac{1}{\sqrt{T}} \frac{1}{\sqrt{T}} \sum_{t=1}^T [\epsilon_{i,t} u'_{i,t} Z_i - \sigma_{\epsilon,i}^2(1,0) Z_i] + \frac{1}{\sqrt{T}} \frac{1}{\sqrt{T}} \sum_{t=1}^T \sum_{j \neq i}^N \epsilon_{i,t} u'_{j,t} Z_j \\ &= \sigma_{\epsilon,i}^2(1,0_m) Z_i + O_p(T^{-1/2}). \end{aligned} \quad (\text{A9})$$

■

Proof of Theorem 1.

Since \bar{H} is $k(m+1) \times r$ with $rk(\bar{H}) = r \leq k(m+1)$, the $r \times r$ matrix $\bar{H}'\bar{H}$ is nonsingular. We may therefore define $\bar{H}^- = \bar{H}(\bar{H}'\bar{H})^{-1}$, such that $\bar{H}'\bar{H}^- = I_r$. The equation for y_i can now be written as

$$y_i = x_i \beta + \hat{F} \bar{H}^- \lambda_i - d \bar{H}^- \lambda_i + \epsilon_i, \quad (\text{A10})$$

where $d = \hat{F} - \bar{H} \bar{H}'$ is as before. The estimator of β is given by

$$\hat{\beta} = \left(\sum_{i=1}^N x'_i M_{\hat{F}} x_i \right)^{-1} \sum_{i=1}^N x'_i M_{\hat{F}} y_i,$$

suggesting that

$$\sqrt{NT}(\hat{\beta} - \beta) = \left(\frac{1}{NT} \sum_{i=1}^N x'_i M_{\hat{F}} x_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^N x'_i M_{\hat{F}} (\epsilon_i - d \bar{H}^- \lambda_i). \quad (\text{A11})$$

We begin by considering the second term in the numerator. Clearly, $M_{\hat{F}} d \bar{H}^- = M_{\hat{F}} (\hat{F} - \bar{H} \bar{H}') \bar{H}^- = -M_{\hat{F}} \bar{H} \bar{H}' \bar{H}^-$, and therefore

$$\begin{aligned} -\frac{1}{\sqrt{NT}} \sum_{i=1}^N x'_i M_{\hat{F}} d \bar{H}^- \lambda_i &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N x'_i M_{\hat{F}} F \lambda_i \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \Lambda_i F' M_{\hat{F}} F \lambda_i + \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta'_i M_{\hat{F}} F \lambda_i \\ &= K_1 + K_2. \end{aligned} \quad (\text{A12})$$

Consider K_1 . Using A^+ to denote the Moore-Penrose inverse A , we define $M_{\bar{H}\bar{H}'} = I_T - \bar{H}\bar{H}'(\bar{H}\bar{H}')^+ \bar{H}\bar{H}'$. It is necessary to use the Moore-Penrose inverse in the definition, since $(\bar{H}\bar{H}')^+$ is $k(m+1) \times k(m+1)$ whereas its rank is r . From

$$\bar{H}\bar{H}' M_{\hat{F}} \bar{H}\bar{H}' = d' M_{\hat{F}} d = d' M_{\bar{H}\bar{H}'} d - d' (M_{\bar{H}\bar{H}'} - M_{\hat{F}}) d,$$

we obtain

$$\begin{aligned}
K_1 &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \Lambda_i F' M_{\hat{F}} F \lambda_i \\
&= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \Lambda_i (\bar{H}^-)' \bar{H} F' M_{\hat{F}} F \bar{H}^- \lambda_i \\
&= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \Lambda_i (\bar{H}^-)' d' M_{F\bar{H}'} d \bar{H}^- \lambda_i + \frac{1}{\sqrt{NT}} \sum_{i=1}^N \Lambda_i (\bar{H}^-)' d' (M_{F\bar{H}'} - M_{\hat{F}}) d \bar{H}^- \lambda_i \\
&= K_{11} + K_{12}.
\end{aligned} \tag{A13}$$

Consider K_{12} , from the definitions of $M_{F\bar{H}'}$ and $M_{\hat{F}}$ he have

$$\begin{aligned}
M_{F\bar{H}'} - M_{\hat{F}} &= d(\hat{F}'\hat{F})^{-1}d' + d(\hat{F}'\hat{F})^{-1}\bar{H}F' \\
&\quad + F\bar{H}'(\hat{F}'\hat{F})^{-1}d' + F\bar{H}'[(\hat{F}'\hat{F})^{-1} - (\bar{H}F'F\bar{H}')^+]\bar{H}F',
\end{aligned}$$

suggesting that

$$\begin{aligned}
d'(M_{F\bar{H}'} - M_{\hat{F}})d &= d'd(\hat{F}'\hat{F})^{-1}d'd + d'd(\hat{F}'\hat{F})^{-1}\bar{H}F'd \\
&\quad + d'F\bar{H}'(\hat{F}'\hat{F})^{-1}d'd + d'F\bar{H}'[(\hat{F}'\hat{F})^{-1} - (\bar{H}F'F\bar{H}')^+]\bar{H}F'd.
\end{aligned} \tag{A14}$$

Consider the fourth term. Using the properties of generalized inverses,

$$(\bar{H}F'F\bar{H}')(\bar{H}F'F\bar{H}')^+(\bar{H}F'F\bar{H}') = (\bar{H}F'F\bar{H}').$$

Multiplying the equation above by $(F'F)^{-1}(\bar{H}'\bar{H})^{-1}\bar{H}'$ from the left and $\bar{H}(\bar{H}'\bar{H})^{-1}(F'F)^{-1}$ from the right,

$$\bar{H}'(\bar{H}F'F\bar{H}')^+\bar{H} = (F'F)^{-1}. \tag{A15}$$

Insertion into $F\bar{H}'[(\hat{F}'\hat{F})^{-1} - (\bar{H}F'F\bar{H}')^+]\bar{H}F'$ yields

$$\begin{aligned}
F\bar{H}'[(\hat{F}'\hat{F})^{-1} - (\bar{H}F'F\bar{H}')^+]\bar{H}F' &= F[\bar{H}'(\hat{F}'\hat{F})^{-1}\bar{H} - (F'F)^{-1}]F' \\
&= F\bar{H}'[(\hat{F}'\hat{F})^{-1} - \bar{H}^-(F'F)^{-1}\bar{H}^-']\bar{H}F' \\
&= F\bar{H}'(\hat{F}'\hat{F})^{-1}[\bar{H}F'F\bar{H}' - \hat{F}'\hat{F}]\bar{H}^-(F'F)^{-1}\bar{H}^-'\bar{H}F' \\
&= F\bar{H}'(\hat{F}'\hat{F})^{-1}[d'F\bar{H}' - \hat{F}'d]\bar{H}^-(F'F)^{-1}\bar{H}^-'\bar{H}F'.
\end{aligned}$$

By Lemmas A.1 and A.2,

$$\begin{aligned}
\|\sqrt{NT}^{-1/2}d'\hat{F}\| &\leq \sqrt{T}N^{-1/2}\|NT^{-1}d'd\| + \|\sqrt{NT}^{-1/2}d'F\|\|\bar{H}\| \\
&= O_p(\sqrt{T}N^{-1/2}) + O_p(1),
\end{aligned} \tag{A16}$$

giving

$$\begin{aligned}
& T \| (\hat{F}' \hat{F})^{-1} - (\bar{H} F' F \bar{H}')^+ \| \\
& \leq \| (T^{-1} \hat{F}' \hat{F})^{-1} \| T^{-1} \| d' \hat{F} + \bar{H} F' d \| \| \bar{H}^- \| \| (T^{-1} F' F)^{-1} \| \| \bar{H}^- \| \\
& = O_p(N^{-1}) + O_p((NT)^{-1/2}),
\end{aligned} \tag{A17}$$

and therefore

$$\begin{aligned}
& \| T^{-1} d' (M_{F\bar{H}'} - M_{\hat{F}}) d \| \\
& \leq \| T^{-1} d' d \|^2 \| (T^{-1} \hat{F}' \hat{F})^{-1} \| + 2 \| \bar{H} \| \| T^{-1} d' d \| \| (T^{-1} \hat{F}' \hat{F})^{-1} \| \| T^{-1} F' d \| \\
& + \| \bar{H} \|^2 \| T^{-1} d' F \|^2 T \| (\hat{F}' \hat{F})^{-1} - (\bar{H} F' F \bar{H}')^+ \| \\
& = O_p(N^{-2}) + O_p(N^{-1}) O_p((NT)^{-1/2}) + [O_p(N^{-1}) + O_p((NT)^{-1/2})] O_p((NT)^{-1}) \\
& = O_p(N^{-2}) + O_p(N^{-3/2} T^{-1/2}).
\end{aligned} \tag{A18}$$

Hence,

$$\begin{aligned}
& \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \Lambda_i (\bar{H}^-)' d' (M_{F\bar{H}'} - M_{\hat{F}}) d \bar{H}^- \lambda_i \right\| \\
& \leq \sqrt{NT} \| \bar{H}^- \|^2 \| T^{-1} d' (M_{F\bar{H}'} - M_{\hat{F}}) d \| \frac{1}{N} \sum_{i=1}^N \| \Lambda_i \| \| \lambda_i \| \\
& = O_p(\sqrt{T} N^{-3/2}) + O_p(N^{-1}).
\end{aligned} \tag{A19}$$

Consider K_{12} . Since $d' M_{F\bar{H}'} d = d' d - d' F \bar{H} (\bar{H} F' F \bar{H}')^+ \bar{H} F' d$, and using the result in (A15), we have $d' M_{F\bar{H}'} d = d' d - d' F (F' F)^{-1} F' d$ with

$$\begin{aligned}
& \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \Lambda_i (\bar{H}^-)' d' F (F' F)^{-1} F' d \bar{H}^- \lambda_i \right\| \\
& \leq \frac{1}{\sqrt{NT}} \| \bar{H}^- \|^2 \| \bar{H} \|^2 \| \sqrt{NT}^{-1/2} d' F \|^2 \| (T^{-1} F' F)^{-1} \| \frac{1}{N} \sum_{i=1}^N \| \Lambda_i \| \| \lambda_i \| \\
& = O_p((NT)^{-1/2}),
\end{aligned}$$

we obtain

$$K_1 = \frac{\sqrt{T}}{\sqrt{N}} \frac{1}{N} \sum_{i=1}^N \Lambda_i (\bar{H}^-)' NT^{-1} d' d \bar{H}^- \lambda_i + O_p(\sqrt{T} N^{-3/2}) + O_p(N^{-1}) + O_p((NT)^{-1/2}).$$

Application of Lemma A.3 now yields

$$K_1 = \sqrt{T} N^{-1/2} B_1 + O_p(\sqrt{T} N^{-3/2}) + O_p(N^{-1/2}), \tag{A20}$$

where, via $\bar{H}^- = \bar{H}(\bar{H}'\bar{H})^{-1}$,

$$B_1 = \frac{1}{N} \sum_{i=1}^N \Lambda_i (\bar{H}^-)' \Theta \bar{H}^- \lambda_i = \frac{1}{N} \sum_{i=1}^N \Lambda_i (\bar{H}'\bar{H})^{-1} \bar{H}' \Theta \bar{H} (\bar{H}'\bar{H})^{-1} \lambda_i. \quad (\text{A21})$$

Next, consider K_2 . By using $M_{F\bar{H}'} F \lambda_i = M_{F\bar{H}'} F \bar{H}' \bar{H}^- \lambda_i = 0$, and then substitution for $(M_{F\bar{H}'} - M_{\hat{F}})$,

$$\begin{aligned} K_2 &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta'_i M_{\hat{F}} F \lambda_i \\ &= -\frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta'_i (M_{F\bar{H}'} - M_{\hat{F}}) F \lambda_i \\ &= -\frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta'_i d(\hat{F}'\hat{F})^{-1} d' F \lambda_i - \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta'_i d(\hat{F}'\hat{F})^{-1} \bar{H} F' F \lambda_i \\ &\quad - \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta'_i F \bar{H}' (\hat{F}'\hat{F})^{-1} d' F \lambda_i - \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta'_i F \bar{H}' [(\hat{F}'\hat{F})^{-1} - (\bar{H} F' F \bar{H}')^+] \bar{H} F' F \lambda_i \\ &= -K_{21} - \dots - K_{24}. \end{aligned}$$

Since $d'_t(\hat{F}'\hat{F})^{-1} d_s$ and $F'_t(\hat{F}'\hat{F})^{-1} d' F \lambda_i$ are just scalars, the orders of K_{21} and K_{23} can be inferred as follows:

$$\begin{aligned} \|K_{21}\| &= \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta'_i d(\hat{F}'\hat{F})^{-1} d' F \lambda_i \right\| \\ &= \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \eta_{i,t} d'_t(\hat{F}'\hat{F})^{-1} \sum_{s=1}^T d_s F'_s \lambda_i \right\| \\ &= \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^T \sum_{s=1}^T d'_t(\hat{F}'\hat{F})^{-1} d_s \sum_{i=1}^N \eta_{i,t} F'_s \lambda_i \right\| \\ &\leq \sqrt{T} \|(T^{-1} \hat{F}'\hat{F})^{-1}\| \frac{1}{T} \sum_{t=1}^T \|d_t\|^2 \left(\frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \eta_{i,t} F'_s \lambda_i \right\|^2 \right)^{1/2} \\ &= O_p(\sqrt{TN}^{-1}), \end{aligned}$$

and

$$\begin{aligned}
\|K_{23}\| &= \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' F \bar{H}' (\hat{F}' \hat{F})^{-1} d' F \lambda_i \right\| \\
&= \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \eta_{i,t} F_t' \bar{H}' (\hat{F}' \hat{F})^{-1} d' F \lambda_i \right\| \\
&= \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^T F_t' \bar{H}' (\hat{F}' \hat{F})^{-1} d' F \sum_{i=1}^N \lambda_i \eta_{i,t} \right\| \\
&\leq \frac{1}{\sqrt{N}} \|\bar{H}\| \left(\frac{1}{T} \sum_{t=1}^T \|F_t\|^2 \right)^{1/2} \|(T^{-1} \hat{F}' \hat{F})^{-1}\| \|\sqrt{N} T^{-1/2} d' F\| \\
&\times \left(\frac{1}{T} \sum_{t=1}^T \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \lambda_i \eta_{i,t} \right\|^2 \right)^{1/2} \\
&= O_p(N^{-1/2}).
\end{aligned}$$

Similarly, since $F_t' \bar{H}' [(\hat{F}' \hat{F})^{-1} - (\bar{H} F' F \bar{H}')^+] \bar{H} F' F \lambda_i$ is a scalar,

$$\begin{aligned}
\|K_{24}\| &= \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' F \bar{H}' T [(\hat{F}' \hat{F})^{-1} - (\bar{H} F' F \bar{H}')^+] T^{-1} \bar{H} F' F \lambda_i \right\| \\
&= \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^T F_t' \bar{H}' [(\hat{F}' \hat{F})^{-1} - (\bar{H} F' F \bar{H}')^+] \bar{H} F' F \sum_{i=1}^N \lambda_i \eta_{i,t} \right\| \\
&\leq \sqrt{T} \|\bar{H}\|^2 \left(\frac{1}{T} \sum_{t=1}^T \|F_t\|^2 \right)^{1/2} T \|(\hat{F}' \hat{F})^{-1} - (\bar{H} F' F \bar{H}')^+\| \|T^{-1} F' F\| \\
&\times \left(\frac{1}{T} \sum_{t=1}^T \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \lambda_i \eta_{i,t} \right\|^2 \right)^{1/2} \\
&= O_p(\sqrt{T} N^{-1}) + O_p(N^{-1/2}).
\end{aligned}$$

K_{22} can be expanded as follows, by adding and subtracting $\bar{H}^- (T^{-1} F' F)^{-1} \bar{H}^{-'}$:

$$\begin{aligned}
K_{22} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' d (\hat{F}' \hat{F})^{-1} \bar{H} F' F \lambda_i \\
&= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' d (T^{-1} \hat{F}' \hat{F})^{-1} T^{-1} \bar{H} F' F \bar{H}' \bar{H}^- \lambda_i \\
&= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' d \bar{H}^- \lambda_i \\
&+ \sqrt{NT} \frac{1}{N} \sum_{i=1}^N T^{-1} \eta_i' d [(T^{-1} \hat{F}' \hat{F})^{-1} - \bar{H}^- (T^{-1} F' F)^{-1} \bar{H}^{-'}] T^{-1} \bar{H} F' F \lambda_i,
\end{aligned}$$

where the summand in last term on the right is

$$\begin{aligned}
& T^{-1} \|\eta'_i d[(T^{-1}\hat{F}'\hat{F})^{-1} - \bar{H}^{-}(T^{-1}F'F)^{-1}\bar{H}^{-'}]T^{-1}\bar{H}F'F\bar{H}'\bar{H}^{-}\lambda_i\| \\
& \leq \left(\frac{1}{T} \sum_{t=1}^T \|\eta_{i,t}\|^2\right)^{1/2} \left(\frac{1}{T} \sum_{t=1}^T \|d_t\|^2\right)^{1/2} \|(T^{-1}\hat{F}'\hat{F})^{-1} - \bar{H}^{-}(T^{-1}F'F)^{-1}\bar{H}^{-'}\| \\
& \times \|\bar{H}^{-}\| \|\lambda_i\| \\
& = O_p(N^{-1/2})[O_p(N^{-1}) + O_p((NT)^{-1/2})] \\
& = O_p(T^{-1/2}N^{-1}) + O_p(N^{-3/2}).
\end{aligned}$$

The order of the second term in K_{22} is \sqrt{NT} times this, which is $O_p(N^{-1/2}) + O_p(\sqrt{TN}^{-1})$.

The first term is, via Lemma A.4,

$$\begin{aligned}
\frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta'_i d\bar{H}^{-}\lambda_i &= \frac{\sqrt{T}}{\sqrt{N}} \frac{1}{T} \sum_{i=1}^N \eta'_i d\bar{H}^{-}\lambda_i \\
&= \frac{\sqrt{T}}{\sqrt{N}} \frac{1}{N} \sum_{i=1}^N \Sigma_{\eta,i}(\beta, I_m) Z_i \bar{H}^{-}\lambda_i + O_p(N^{-1/2})
\end{aligned}$$

By using this we obtain

$$K_{22} = \sqrt{TN}^{-1/2} B_2 + O_p(\sqrt{TN}^{-1}) + O_p(N^{-1/2}), \quad (\text{A22})$$

where $B_2 = N^{-1} \sum_{i=1}^N \Sigma_{\eta,i}(\beta, I_m) Z_i \bar{H}^{-}(\bar{H}'\bar{H})^{-1}\lambda_i$. Thus, by adding the results,

$$\begin{aligned}
-\frac{1}{\sqrt{NT}} \sum_{i=1}^N x'_i M_{\hat{F}} d\bar{H}^{-}\lambda_i &= K_1 + K_2 \\
&= \sqrt{TN}^{-1/2} (B_1 - B_2) + O_p(N^{-1/2}) + O_p(\sqrt{TN}^{-1}). \quad (\text{A23})
\end{aligned}$$

Next, consider $(NT)^{-1/2} \sum_{i=1}^N x'_i M_{\hat{F}} \epsilon_i$, the first term in the numerator of $\sqrt{NT}(\hat{\beta} - \beta)$.

Clearly,

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^N x'_i M_{\hat{F}} \epsilon_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^N x'_i M_{F\bar{H}'} \epsilon_i - \frac{1}{\sqrt{NT}} \sum_{i=1}^N x'_i (M_{F\bar{H}'} - M_{\hat{F}}) \epsilon_i, \quad (\text{A24})$$

where

$$\begin{aligned}
& \frac{1}{\sqrt{NT}} \sum_{i=1}^N x'_i (M_{F\bar{H}'} - M_{\hat{F}}) \epsilon_i \\
&= \frac{1}{\sqrt{NT}} \sum_{i=1}^N x'_i d(\hat{F}'\hat{F})^{-1} d' \epsilon_i + \frac{1}{\sqrt{NT}} \sum_{i=1}^N x'_i d(\hat{F}'\hat{F})^{-1} \bar{H}F' \epsilon_i \\
&+ \frac{1}{\sqrt{NT}} \sum_{i=1}^N x'_i F\bar{H}' (\hat{F}'\hat{F})^{-1} d' \epsilon_i + \frac{1}{\sqrt{NT}} \sum_{i=1}^N x'_i F\bar{H}' [(\hat{F}'\hat{F})^{-1} - (\bar{H}F'F\bar{H}')^+] \bar{H}F' \epsilon_i \\
&= L_1 + \dots + L_4. \quad (\text{A25})
\end{aligned}$$

The order of L_1, \dots, L_4 can be obtained by using the same steps as when analyzing K_2 . For L_1 , we use the fact that $x_{i,t} = \Lambda_i F_t + \eta_{i,t}$, giving

$$\left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N x_{i,t} \epsilon_{i,s} \right\| \leq \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \Lambda_i \epsilon_{i,s} \right\| \|F_t\| + \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \eta_{i,t} \epsilon_{i,s} \right\| = O_p(1),$$

which in turn implies

$$\begin{aligned} \|L_1\| &= \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N x'_i d (\hat{F}' \hat{F})^{-1} d' \epsilon_i \right\| \\ &= \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^T \sum_{s=1}^T d'_t (\hat{F}' \hat{F})^{-1} d_s \sum_{i=1}^N x_{i,t} \epsilon_{i,s} \right\| \\ &\leq \sqrt{T} \frac{1}{T} \sum_{t=1}^T \|d_t\|^2 \|(T^{-1} \hat{F}' \hat{F})^{-1}\| \left(\frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N x_{i,t} \epsilon_{i,s} \right\|^2 \right)^{1/2} \\ &= O_p(\sqrt{T} N^{-1}). \end{aligned} \tag{A26}$$

Similarly,

$$\begin{aligned} \|L_4\| &= \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N x'_i F \bar{H}' [(\hat{F}' \hat{F})^{-1} - (\bar{H} F' F \bar{H}')^+] \bar{H} F' \epsilon_i \right\| \\ &\leq \sqrt{N} \|\bar{H}\|^2 \left(\frac{1}{N} \sum_{i=1}^N \|T^{-1} x'_i F\|^2 \right)^{1/2} T \|(\hat{F}' \hat{F})^{-1} - (\bar{H} F' F \bar{H}')^+\| \\ &\times \left(\frac{1}{N} \sum_{i=1}^N \|T^{-1/2} F' \epsilon_i\|^2 \right)^{1/2} \\ &= O_p(N^{-1/2}) + O_p(T^{-1/2}). \end{aligned} \tag{A27}$$

Consider L_2 . Adding and subtracting $\bar{H}^- (F' F)^{-1} \bar{H}^{-'}$ give

$$\begin{aligned} L_2 &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N x'_i d (\hat{F}' \hat{F})^{-1} \bar{H} F' \epsilon_i \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N x'_i d \bar{H}^- (F' F)^{-1} F' \epsilon_i + \frac{1}{\sqrt{NT}} \sum_{i=1}^N x'_i d [(\hat{F}' \hat{F})^{-1} - \bar{H}^- (F' F)^{-1} \bar{H}^{-'}] \bar{H} F' \epsilon_i \\ &= L_{21} + L_{22}, \end{aligned}$$

where

$$\begin{aligned}
\|L_{22}\| &= \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N x_i' d [(\hat{F}' \hat{F})^{-1} - \bar{H}^{-} (F' F)^{-1} \bar{H}^{-'}] \bar{H} F' \epsilon_i \right\| \\
&= \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T x_{i,t} d_t' [(\hat{F}' \hat{F})^{-1} - \bar{H}^{-} (F' F)^{-1} \bar{H}^{-'}] \bar{H} F_s \epsilon_{i,s} \right\| \\
&= \sqrt{T} \|\bar{H}\| \left(\frac{1}{T} \sum_{t=1}^T \|d_t\|^2 \right)^{1/2} T \|(\hat{F}' \hat{F})^{-1} - \bar{H}^{-} (F' F)^{-1} \bar{H}^{-'}\| \\
&\quad \times \left(\frac{1}{T} \sum_{s=1}^T \|F_s\|^2 \right)^{1/2} \left(\frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \epsilon_{i,s} x_{i,t} \right\|^2 \right)^{1/2} \\
&= \sqrt{T} O_p(N^{-1/2}) [O_p(N^{-1}) + O_p((NT)^{-1/2})] \\
&= O_p(N^{-1}) + O_p(\sqrt{T} N^{-3/2}).
\end{aligned}$$

Also, from $x_i = F \Lambda_i' + \eta_i$,

$$\begin{aligned}
L_{21} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N x_i' d \bar{H}^{-} (F' F)^{-1} F' \epsilon_i \\
&= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \Lambda_i F' d \bar{H}^{-} (F' F)^{-1} F' \epsilon_i + \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' d \bar{H}^{-} (F' F)^{-1} F' \epsilon_i.
\end{aligned}$$

Lemma A.4 implies $\|NT^{-1} \eta_i' d\| = O_p(1)$, from which it follows that

$$\begin{aligned}
&\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' d \bar{H}^{-} (F' F)^{-1} F' \epsilon_i \right\| \\
&\leq \frac{1}{\sqrt{N}} \frac{1}{N} \sum_{i=1}^N \|NT^{-1} \eta_i' d\| \|\bar{H}^{-}\| \|(T^{-1} F' F)^{-1}\| \|T^{-1/2} F' \epsilon_i\| = O_p(N^{-1/2}),
\end{aligned}$$

and by further use of Lemma A.2,

$$\begin{aligned}
&\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \Lambda_i F' d \bar{H}^{-} (F' F)^{-1} F' \epsilon_i \right\| \\
&\leq \frac{1}{\sqrt{T}} \frac{1}{N} \sum_{i=1}^N \|\Lambda_i\| \|\sqrt{N} T^{-1/2} F' d\| \|\bar{H}^{-}\| \|(T^{-1} F' F)^{-1}\| \|T^{-1/2} F' \epsilon_i\| \\
&= O_p(T^{-1/2}).
\end{aligned}$$

Consequently,

$$\begin{aligned}
\|L_{21}\| &\leq \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \Lambda_i F' d \bar{H}^{-} (F' F)^{-1} F' \epsilon_i \right\| + \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' d \bar{H}^{-} (F' F)^{-1} F' \epsilon_i \right\| \\
&= O_p(N^{-1/2}) + O_p(T^{-1/2}),
\end{aligned}$$

leading to the following result for $\|L_2\|$:

$$\|L_2\| \leq \|L_{21}\| + \|L_{22}\| = O_p(N^{-1/2}) + O_p(T^{-1/2}) + O_p(\sqrt{T}N^{-3/2}). \quad (\text{A28})$$

Consider L_3 . As when evaluating L_2 , we begin by adding and subtracting $\bar{H}^- (F'F)^{-1} \bar{H}^{-'}$;

$$\begin{aligned} L_3 &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N x_i' F \bar{H}' (\hat{F}' \hat{F})^{-1} d' \epsilon_i \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N x_i' F (F'F)^{-1} \bar{H}^{-'} d' \epsilon_i + \frac{1}{\sqrt{NT}} \sum_{i=1}^N x_i' F \bar{H}' [(\hat{F}' \hat{F})^{-1} - \bar{H}^- (F'F)^{-1} \bar{H}^{-'}] d' \epsilon_i, \end{aligned}$$

where, in analogy to $\|L_{22}\|$,

$$\begin{aligned} &\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N x_i' F \bar{H}' [(\hat{F}' \hat{F})^{-1} - \bar{H}^- (F'F)^{-1} \bar{H}^{-'}] d' \epsilon_i \right\| \\ &= \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T x_{i,t} F_t' \bar{H}' [(\hat{F}' \hat{F})^{-1} - \bar{H}^- (F'F)^{-1} \bar{H}^{-'}] d_s \epsilon_{i,s} \right\| \\ &\leq \sqrt{T} \left(\frac{1}{T} \sum_{t=1}^T \|F_t\|^2 \right)^{1/2} \|\bar{H}\| T \|(\hat{F}' \hat{F})^{-1} - \bar{H}^- (F'F)^{-1} \bar{H}^{-'}\| \left(\frac{1}{T} \sum_{s=1}^T \|d_s\|^2 \right)^{1/2} \\ &\times \left(\frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N x_{i,t} \epsilon_{i,s} \right\|^2 \right)^{1/2} \\ &= \sqrt{T} [O_p(N^{-1}) + O_p((NT)^{-1/2})] O_p(N^{-1/2}) \\ &= O_p(N^{-1}) + O_p(\sqrt{T}N^{-3/2}). \end{aligned}$$

and, by substituting for x_i , via Lemma A.5 the first term in L_3 can be written

$$\begin{aligned} &\frac{1}{\sqrt{NT}} \sum_{i=1}^N x_i' F (F'F)^{-1} \bar{H}^{-'} d' \epsilon_i \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \Lambda_i F' F (F'F)^{-1} \bar{H}^{-'} d' \epsilon_i + \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' F (F'F)^{-1} \bar{H}^{-'} d' \epsilon_i \\ &= \frac{\sqrt{T}}{\sqrt{N}} \frac{1}{N} \sum_{i=1}^N \Lambda_i (\bar{H}^-)' N T^{-1} d' \epsilon_i + \frac{1}{\sqrt{N}} \frac{1}{N} \sum_{i=1}^N T^{-1/2} \eta_i' F (F'F)^{-1} \bar{H}^{-'} d' \epsilon_i \\ &= \frac{\sqrt{T}}{\sqrt{N}} \frac{1}{N} \sum_{i=1}^N \sigma_{\epsilon,i}^2 \Lambda_i (\bar{H}^-)' Z_i' (1, 0_m)' + O_p(N^{-1/2}). \end{aligned}$$

Hence, letting $B_3 = N^{-1} \sum_{i=1}^N \sigma_{\epsilon,i}^2 \Lambda_i (\bar{H}^-)' Z_i' (1, 0_m)'$, we obtain

$$L_3 = \sqrt{T} N^{-1/2} B_3 + O_p(N^{-1/2}) + O_p(\sqrt{T}N^{-3/2}). \quad (\text{A29})$$

The results for L_1, \dots, L_4 imply

$$\begin{aligned}
& \frac{1}{\sqrt{NT}} \sum_{i=1}^N x'_i (M_{F\bar{H}'} - M_{\hat{F}}) \epsilon_i \\
&= L_1 + \dots + L_4 \\
&= \sqrt{T} N^{-1/2} B_3 + O_p(\sqrt{T} N^{-1}) + O_p(T^{-1/2}) + O_p(N^{-1/2}).
\end{aligned} \tag{A30}$$

which in turn implies

$$\begin{aligned}
\frac{1}{\sqrt{NT}} \sum_{i=1}^N x'_i M_{\hat{F}} \epsilon_i &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N x'_i M_{F\bar{H}'} \epsilon_i - \sqrt{T} N^{-1/2} B_3 + O_p(\sqrt{T} N^{-1}) \\
&+ O_p(T^{-1/2}) + O_p(N^{-1/2}).
\end{aligned} \tag{A31}$$

Consider the first term in (A44). By using the cross-section independence of ϵ_i , the serial independence of $\eta_{i,t}$, $E(\epsilon_i \epsilon_i') = \sigma_{\epsilon,i}^2 I_T$ and $E(\eta_{i,t} \eta_{i,t}') = \Sigma_{\eta,i}$ and the result in (A15), we can show that

$$\begin{aligned}
& E \left[\left(\frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' F \bar{H}' (\bar{H} F' F \bar{H}')^+ \bar{H} F' \epsilon_i \right) \left(\frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' F \bar{H}' (\bar{H} F' F \bar{H}')^+ \bar{H} F' \epsilon_i \right)' \right] \\
&= \frac{1}{NT} \sum_{i=1}^N \sum_{j=1}^N E[\eta_i' F (F' F)^{-1} F' E(\epsilon_i \epsilon_j' | \eta_i, \eta_j, F) F (F' F)^{-1} F' \eta_j] \\
&= \frac{1}{NT} \sum_{i=1}^N \sigma_{\epsilon,i}^2 E[\eta_i' F (F' F)^{-1} F' F (F' F)^{-1} F' \eta_i] \\
&= \frac{1}{NT} \sum_{i=1}^N \sigma_{\epsilon,i}^2 E[\eta_i' F (F' F)^{-1} F' \eta_i] \\
&= \frac{1}{NT} \sum_{i=1}^N \sigma_{\epsilon,i}^2 \sum_{t=1}^T \sum_{s=1}^T E[F_t' (F' F)^{-1} F_s E(\eta_{i,t} \eta_{i,s}' | F)] \\
&= \frac{1}{NT} \sum_{i=1}^N \sigma_{\epsilon,i}^2 \sum_{t=1}^T E[F_t' (F' F)^{-1} F_t \Sigma_{\eta,i}] \\
&= \frac{1}{T} \frac{r}{N} \sum_{i=1}^N \sigma_{\epsilon,i}^2 \Sigma_{\eta,i} = O_p(T^{-1}),
\end{aligned}$$

where last equality holds because

$$\sum_{t=1}^T F_t' (F' F)^{-1} F_t = \sum_{t=1}^T \text{tr}(F_t' (F' F)^{-1} F_t) = \text{tr} \left(\sum_{t=1}^T F_t F_t' (F' F)^{-1} \right) = \text{tr}(I_r) = r.$$

Thus, since the variance is $O_p(T^{-1})$, we have

$$\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' F \bar{H}' (\bar{H} F' F \bar{H}')^+ \bar{H} F' \epsilon_i \right\| = O_p(T^{-1/2}). \tag{A32}$$

This result, together with the fact that $M_{F\bar{H}}x_i = M_{F\bar{H}}\eta_i$, implies

$$\begin{aligned}
\frac{1}{\sqrt{NT}} \sum_{i=1}^N x_i' M_{F\bar{H}} \epsilon_i &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' M_{F\bar{H}} \epsilon_i \\
&= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' \epsilon_i - \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' F\bar{H}(\bar{H}' F' F\bar{H})^+ \bar{H}' F' \epsilon_i \\
&= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' \epsilon_i + O_p(T^{-1/2}).
\end{aligned} \tag{A33}$$

Consider the variance of $\frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' \epsilon_i$. We have

$$\begin{aligned}
E(\epsilon_{i,t} \epsilon_{i,t-s}) &= E[\phi_i(L) \epsilon_{i,t} \phi_i(L) \epsilon_{i,t-s}] = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \phi_{j,i} \phi_{k,i} E(\epsilon_{i,t-j} \epsilon_{i,t-s-k}) \\
&= \sum_{k=0}^{\infty} \phi_{s+k,i} \phi_{k,i} E(\epsilon_{i,t-s-k}^2) = \sum_{k=0}^{\infty} \phi_{s+k,i} \phi_{k,i} = \gamma_{s,i}, \\
E(\eta_{i,t} \eta_{i,t-s}') &= E[\Phi_i(L) v_{i,t} (\phi_i(L) v_{i,t-s})'] = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \Phi_{j,i} E(v_{i,t-j} v_{i,t-s-k}') \Phi_{k,i}' \\
&= \sum_{k=0}^{\infty} \Phi_{s+k,i} E(v_{i,t-s-k} v_{i,t-s-k}') \Phi_{k,i}' = \sum_{k=0}^{\infty} \Phi_{s+k,i} \Phi_{k,i}' = \Gamma_{s,i}
\end{aligned}$$

for any $s = -(T-t), \dots, t-1$. Hence, letting $W_i = T^{-1} \sum_{t=1}^T \sum_{s=-(T-t)}^{t-1} \gamma_{s,i} \Gamma_{s,i}$ and $\bar{W} = N^{-1} \sum_{i=1}^N W_i$,

$$\begin{aligned}
&E \left[\left(\frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' \epsilon_i \right) \left(\frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' \epsilon_i \right)' \right] \\
&= \frac{1}{NT} \sum_{i=1}^N \sum_{j=1}^N E[\eta_i' E(\epsilon_i \epsilon_j' | \eta_i, \eta_j) \eta_j] = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T E(\epsilon_{i,t} \epsilon_{i,s}) E(\eta_{i,t} \eta_{i,s}') \\
&= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T E(\epsilon_{i,t}^2) E(\eta_{i,t} \eta_{i,t}') + \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^{t-1} E(\epsilon_{i,t} \epsilon_{i,t-s}) E(\eta_{i,t} \eta_{i,t-s}') \\
&+ \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^{T-t} E(\epsilon_{i,t} \epsilon_{i,t+s}) E(\eta_{i,t} \eta_{i,t+s}') \\
&= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \gamma_{0,i} \Gamma_{0,i} + \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^{t-1} \gamma_{s,i} \Gamma_{s,i} + \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^{T-t} \gamma_{-s,i} \Gamma_{-s,i} \\
&= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=-(T-t)}^{t-1} \gamma_{s,i} \Gamma_{s,i} = \bar{W}.
\end{aligned}$$

Moreover, since the forth-order moments of $\eta_{i,t}$ and $\epsilon_{i,t}$ are bounded by assumption, by a central limit law for iid variates,

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' \epsilon_i \sim N(0, \bar{W}) \tag{A34}$$

as $N, T \rightarrow \infty$, where \sim signifies asymptotic equivalence. Thus, provided that $\sqrt{TN}^{-1} = o(1)$,

$$\begin{aligned} \frac{1}{\sqrt{NT}} \sum_{i=1}^N x'_i M_{\hat{F}} \epsilon_i &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta'_i \epsilon_i - \sqrt{TN}^{-1/2} B_3 + O_p(\sqrt{TN}^{-1}) \\ &+ O_p(T^{-1/2}) + O_p(N^{-1/2}) \\ &\sim N(0, \bar{W}) - \sqrt{TN}^{-1/2} B_3. \end{aligned} \quad (\text{A35})$$

Let $B = B_1 - B_2 - B_3$. The above results suggest the following limit for the numerator of $\sqrt{NT}(\hat{\beta} - \beta)$:

$$\begin{aligned} \frac{1}{\sqrt{NT}} \sum_{i=1}^N (x'_i M_{\hat{F}} \epsilon_i - x'_i M_{\hat{F}} d \bar{H}^- \lambda_i) \\ = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta'_i \epsilon_i + \sqrt{TN}^{-1/2} B + O_p(\sqrt{TN}^{-1}) + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\ \sim N(0, \bar{W}) + \sqrt{TN}^{-1/2} B, \end{aligned} \quad (\text{A36})$$

which holds as $N, T \rightarrow \infty$ with $\sqrt{TN}^{-1} \rightarrow 0$.

Next, consider the denominator of $\sqrt{NT}(\hat{\beta} - \beta)$, which we expand as

$$\frac{1}{NT} \sum_{i=1}^N x'_i M_{\hat{F}} x_i = \frac{1}{NT} \sum_{i=1}^N x'_i M_{F\bar{H}'} x_i - \frac{1}{NT} \sum_{i=1}^N x'_i (M_{F\bar{H}'} - M_{\hat{F}}) x_i. \quad (\text{A37})$$

By Lemma A.3, $NT^{-1} \|d\|^2 = O_p(1)$, implying $\sqrt{NT}^{-1/2} \|d\| = O_p(1)$. Similarly, since $T^{-1} \|F\|^2 = T^{-1} \sum_{t=1}^T \|F_t\|^2 = O_p(1)$, we have $T^{-1/2} \|F\| = O_p(1)$. Hence,

$$\begin{aligned} \|M_{F\bar{H}'} - M_{\hat{F}}\| &\leq T^{-1} \|d\|^2 \|(T^{-1} \hat{F}' \hat{F})^{-1}\| + 2T^{-1} \|\bar{H}\| \|d\| \|F\| \|(T^{-1} \hat{F}' \hat{F})^{-1}\| \\ &+ T^{-1} \|\bar{H}\|^2 \|F\|^2 T \|(\hat{F}' \hat{F})^{-1} - (\bar{H} F' F \bar{H}')^+\| \\ &= O_p(N^{-1/2}). \end{aligned}$$

implying

$$\left\| \frac{1}{NT} \sum_{i=1}^N x'_i (M_{F\bar{H}'} - M_{\hat{F}}) x_i \right\| \leq \|M_{F\bar{H}'} - M_{\hat{F}}\| \frac{1}{NT} \sum_{i=1}^N \|x_i\|^2 = O_p(N^{-1/2}).$$

By using this and

$$x'_i M_{F\bar{H}'} x_i = \eta'_i M_{F\bar{H}'} \eta_i = \eta'_i \eta_i - T^{-1/2} \eta'_i F (T^{-1} F' F)^{-1} T^{-1/2} F' \eta_i = \eta'_i \eta_i + O_p(1),$$

we obtain

$$\begin{aligned}
\frac{1}{NT} \sum_{i=1}^N x_i' M_{\hat{F}} x_i &= \frac{1}{NT} \sum_{i=1}^N x_i' M_{F\bar{H}'} x_i + O_p(N^{-1/2}) \\
&= \bar{\Sigma}_\eta + \frac{1}{N} \sum_{i=1}^N (T^{-1} \eta_i' \eta_i - \Sigma_{\eta,i}) + O_p(N^{-1/2}) \\
&= \bar{\Sigma}_\eta + O_p(N^{-1/2}).
\end{aligned} \tag{A38}$$

By adding all the results, as $N, T \rightarrow \infty$ with $\sqrt{TN}^{-1} \rightarrow 0$,

$$\begin{aligned}
&\sqrt{NT}(\hat{\beta} - \beta) \\
&= \left(\frac{1}{NT} \sum_{i=1}^N x_i' M_{\hat{F}} x_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^N (x_i' M_{\hat{F}} \epsilon_i - x_i' M_{\hat{F}} d\bar{H}^{-1} \lambda_i) \\
&= \bar{\Sigma}_\eta^{-1} \left(\frac{1}{\sqrt{NT}} \sum_{i=1}^N \eta_i' \epsilon_i + \sqrt{T} N^{-1/2} B \right) + O_p(\sqrt{TN}^{-1}) + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\
&\sim N(0, \bar{\Sigma}_\eta^{-1} \bar{W} \bar{\Sigma}_\eta^{-1}) + \bar{\Sigma}_\eta^{-1} \sqrt{T} N^{-1/2} B.
\end{aligned}$$

This completes the proof. ■

Proof of Proposition 1.

Write

$$\begin{aligned}
\sqrt{NT}(\hat{\beta}_{BA} - \beta) &= \sqrt{NT}(\hat{\beta} - \beta) - \sqrt{T} N^{-1/2} \hat{\Sigma}_\eta^{-1} \hat{B} \\
&= \sqrt{NT}(\hat{\beta} - \beta) - \sqrt{T} N^{-1/2} \Sigma_\eta^{-1} B - \sqrt{T} N^{-1/2} \hat{\Sigma}_\eta^{-1} (\hat{B} - B) \\
&\quad - \sqrt{T} N^{-1/2} (\hat{\Sigma}_\eta^{-1} - \Sigma_\eta^{-1}) B.
\end{aligned} \tag{A39}$$

Consider $\sqrt{T} N^{-1/2} \hat{\Sigma}_\eta^{-1} (\hat{B} - B)$. From Theorem 1, we have that $(\hat{\beta} - \beta) = O_p((NT)^{-1/2}) + O_p(N^{-1})$. \hat{C}_i is given by

$$\hat{C}_i = (T^{-1} \hat{F}' \hat{F})^{-1} T^{-1} \hat{F}' w_i,$$

where

$$\begin{aligned}
T^{-1} \hat{F}' w_i &= T^{-1} \hat{F}' F C_i + T^{-1} \hat{F}' u_i \\
&= T^{-1} \bar{H} F' F C_i + T^{-1} d' F C_i + T^{-1} \hat{F}' u_i \\
&= T^{-1} \bar{H} F' F C_i + T^{-1} \bar{H} F' u_i + T^{-1} d' F C_i + T^{-1} d' u_i.
\end{aligned}$$

Clearly, $\|T^{-1} F' u_i\| = O_p(T^{-1/2})$, and by Lemmas A.2, A.4 and A.5, $\|T^{-1} d' F\|$ and $\|T^{-1} d' u_i\|$ are $O_p((NT)^{-1/2})$. Moreover, from the proof of Theorem 1,

$$T \|(\hat{F}' \hat{F})^{-1} - \bar{H}^{-1} (F' F)^{-1} \bar{H}^{-1}\| = O_p(N^{-1}) + O_p((NT)^{-1/2}).$$

It follows that, with $\bar{H}^- = \bar{H}(\bar{H}'\bar{H})^{-1}$,

$$\begin{aligned}
\hat{C}_i &= (T^{-1}\hat{F}'\hat{F})^{-1}T^{-1}\hat{F}'w_i \\
&= \bar{H}^-(T^{-1}F'F)^{-1}\bar{H}'(T^{-1}\bar{H}F'FC_i + T^{-1}\bar{H}F'u_i + T^{-1}d'FC_i + T^{-1}d'u_i) + O_p(N^{-1}) \\
&\quad + O_p((NT)^{-1/2}) \\
&= \bar{H}^-(T^{-1}F'F)^{-1}\bar{H}'T^{-1}\bar{H}F'FC_i + O_p(N^{-1}) + O_p(T^{-1/2}) \\
&= \bar{H}^-C_i + O_p(N^{-1}) + O_p(T^{-1/2})
\end{aligned} \tag{A40}$$

$(\hat{\Sigma}_\eta - \Sigma_\eta)$, $(\hat{\Sigma}_{\eta,i} - \Sigma_{\eta,i})$ and $(\hat{\sigma}_{\epsilon,i}^2 - \sigma_{\epsilon,i}^2)$ are all $O_p(T^{-1/2})$ (details are available upon request).

This implies that

$$\|\hat{B} - B\| = O_p(T^{-1/2}) + O_p(N^{-1}),$$

and therefore, with $\|\hat{\Sigma}_\eta^{-1}\| = O_p(1)$,

$$\|\sqrt{TN}^{-1/2}\hat{\Sigma}_\eta^{-1}(\hat{B} - B)\| \leq \sqrt{TN}^{-1/2}\|\hat{\Sigma}_\eta^{-1}\|\|\hat{B} - B\| = O_p(N^{-1/2}) + O_p(\sqrt{TN}^{-3/2}), \tag{A41}$$

which is $o_p(1)$ under our assumption that $\sqrt{TN}^{-1} = o(1)$. Similarly, since $\|B\| = O_p(1)$

and, by Taylor expansion, $\|\hat{\Sigma}_\eta^{-1} - \Sigma_\eta^{-1}\| = O_p(T^{-1/2})$,

$$\|\sqrt{TN}^{-1/2}(\hat{\Sigma}_\eta^{-1} - \Sigma_\eta^{-1})B\| \leq \sqrt{TN}^{-1/2}\|\hat{\Sigma}_\eta^{-1} - \Sigma_\eta^{-1}\|\|B\| = O_p(N^{-1/2}). \tag{A42}$$

Together with Theorem 1 these results imply

$$\sqrt{NT}(\hat{\beta}_{BA} - \beta) = \sqrt{NT}(\hat{\beta} - \beta) - \sqrt{\tau}\Sigma_\eta^{-1}B + o_p(1) \rightarrow_d N(0, \bar{\Sigma}_\eta^{-1}\bar{W}\bar{\Sigma}_\eta^{-1})$$

as $N, T \rightarrow \infty$ with $\sqrt{TN}^{-1} \rightarrow 0$ and $\sqrt{NT}^{-1} \rightarrow 0$. ■

Proof of Proposition 2.

Theorem 2 of Bai and Ng (2002) states the consistency of the IC when applied to the PC estimator. According to Corollary 2 of Bai and Ng (2002) this result holds also when the method of estimation for the factors differs from PC, provided that the rate of consistency of the factor estimator is not greater than the rate that applies to PC, which is $\min\{\sqrt{N}, \sqrt{T}\}$. In our case, $\|d_t\| = O_p(N^{-1/2})$, which means that the condition of Corollary 2 is not satisfied when $N < T$. Hence, since we cannot rely on Corollary 2 of Bai and Ng (2002), we adapt the proof of their Theorem 2 to the current setup.

We need to prove that

$$P[IC(s) < IC(k)] = P[V(s, \hat{F}^s) - V(k, \hat{F}^k) < (k - s)g(N)] \rightarrow 0 \tag{A43}$$

as $N, T \rightarrow \infty$. Consider first the case when $s < k$. Under the condition that $g(N) \rightarrow 0$ as $N \rightarrow \infty$, it is sufficient to show that $V(s, \hat{F}^s) - V(k, \hat{F}^k)$ has a positive limit for each $1 \leq s < k$.

For this purpose, we use the identity

$$\begin{aligned} V(s, \hat{F}^s) - V(k, \hat{F}^k) &= [V(s, \hat{F}^s) - V(s, F\bar{H}^s)] + [V(s, F\bar{H}^s) - V(k, F\bar{H}^k)] \\ &\quad + [V(k, F\bar{H}^k) - V(k, \hat{F}^k)]. \end{aligned} \quad (\text{A44})$$

It can be established that

$$V(s, \hat{F}^s) - V(s, F\bar{H}^s) = O_p(N^{-1/2}).$$

The first and the third terms of (A44) are of this form, so they are $O_p(N^{-1/2})$. The proof of this result is completely analogous to the proof of Lemma 2 in Bai and Ng (2002). For the second term of (A44), we can show that, for any s with $1 \leq s \leq k$, there exists a $\kappa_s > 0$, such that

$$V(s, F\bar{H}^s) - V(k, F\bar{H}^k) \rightarrow \kappa_s$$

as $N, T \rightarrow \infty$. The proof of this result is completely analogous to the proof of Lemma 3 in Bai and Ng (2002). It follows that, if $s < k$, then $V(s, \hat{F}^s) - V(k, \hat{F}^k) = \kappa_s + o_p(1)$, as required.

Next, consider the case when $s \geq k$. The rank condition in Assumption 2 is satisfied in this case. Using this we now show that

$$|V(k, \hat{F}^k) - V(s, \hat{F}^s)| = O_p(N^{-1}).$$

We begin by noting that when $s \geq k$,

$$|V(k, \hat{F}^k) - V(s, \hat{F}^s)| \leq 2 \max_{k \leq s \leq k_{\max}} |V(k, F) - V(s, \hat{F}^s)|.$$

Hence, if we can show that $|V(k, F) - V(s, \hat{F}^s)| = O_p(N^{-1})$, then we are done. Considering (A10) and assuming that $\beta = 0$ to simplify things, we have

$$\begin{aligned} V(s, \hat{F}^s) - V(k, F) &= \frac{1}{NT} \sum_{i=1}^N (\epsilon_i - d\bar{H}^- \lambda_i)' M_{\hat{F}^s} (\epsilon_i - d\bar{H}^- \lambda_i) - \frac{1}{NT} \sum_{i=1}^N \epsilon_i' M_F \epsilon_i \\ &= \frac{1}{NT} \sum_{i=1}^N \epsilon_i' (M_{\hat{F}^s} - M_F) \epsilon_i - \frac{1}{NT} \sum_{i=1}^N \lambda_i' \bar{H}^{-'} d' M_{\hat{F}^s} \epsilon_i \\ &\quad - \frac{1}{NT} \sum_{i=1}^N \epsilon_i' M_{\hat{F}^s} d\bar{H}^- \lambda_i + \frac{1}{NT} \sum_{i=1}^N \lambda_i' \bar{H}^{-'} d' M_{\hat{F}^s} d\bar{H}^- \lambda_i. \end{aligned}$$

Since the rank condition is satisfied, $M_F = M_{F\bar{H}^{-'}}$. The fourth term on the right-hand side of the above expansion is $(NT)^{-1/2}$ times the order of K_1 in the proof of Theorem 1. Hence,

this term is $O_p(N^{-1})$, and so are second and the third terms. To find the order of the first term we write

$$\begin{aligned}\epsilon'_i(M_{\hat{F}} - M_F)\epsilon_i &= \epsilon'_i d(\hat{F}^{s'} \hat{F}^s)^{-1} d' \epsilon_i + \epsilon'_i d(\hat{F}^{s'} \hat{F}^s)^{-1} \bar{H}^s F' \epsilon_i \\ &+ \epsilon'_i F \bar{H}^{s'} (\hat{F}^{s'} \hat{F}^s)^{-1} d' \epsilon_i + \epsilon'_i F \bar{H}^{s'} [(\hat{F}^{s'} \hat{F}^s)^{-1} - (\bar{H}^s F' F \bar{H}^{s'})^+] \bar{H}^s F' \epsilon_i.\end{aligned}$$

Using Lemma A.5, we know that $\epsilon'_i d = O_p(TN^{-1})$, and it is straightforward to show that $F' \epsilon_i = O_p(\sqrt{T})$. These results imply

$$\frac{1}{NT} \sum_{i=1}^N \epsilon'_i (M_{\hat{F}^s} - M_F) \epsilon_i = O_p(N^{-1}),$$

Using this result and the condition that $Ng(N) \rightarrow \infty$ it is obvious that (A43) holds for the case with $s \geq k$. This completes the proof. ■

A Monte Carlo results

Table A: Description of experiments.

Experiment	r	Observables	Factor loadings	Combinations
E1	2	$y_{i,t} = \beta x_{i,t} + \lambda_{1i}f_{1t} + \lambda_{2i}f_{2t} + \epsilon_{i,t}$ $x_{i,t} = \Lambda_{1i}f_{1t} + \Lambda_{2i}f_{2t} + \eta_{i,t}$	$\lambda_{1i} = z_{1i} + z_{2i} + \tau_{\lambda 1i}$ $\lambda_{2i} = z_{1i} + z_{2i} + \tau_{\lambda 2i}$ $\Lambda_{1i} = z_{1i} + \tau_{\Lambda 1i}$ $\Lambda_{2i} = z_{2i} + \tau_{\Lambda 2i}$	$z_{1i} \sim N(1.0, 0.5)$ $z_{2i} \sim N(0.2, 0.5)$
E2	2	$y_{i,t} = \beta x_{i,t} + \lambda_{1i}f_{1t} + \lambda_{2i}f_{2t} + \epsilon_{i,t}$ $x_{i,t} = \Lambda_{1i}f_{1t} + \Lambda_{2i}f_{2t} + \eta_{i,t}$	$\lambda_{1i} = z_{1i} + z_{2i} + \tau_{\lambda 1i}$ $\lambda_{2i} = z_{1i} + z_{3i} + \tau_{\lambda 2i}$ $\Lambda_{1i} = z_{1i} + z_{3i} + \tau_{\Lambda 1i}$ $\Lambda_{2i} = z_{4i} + \tau_{\Lambda 2i}$	$z_{1i} \sim N(1.0, 0.5)$ $z_{2i} \sim N(0.2, 0.5)$ $z_{3i} \sim N(0.5, 0.5)$ $z_{4i} \sim N(0.5, 0.5)$
E3	2	$y_{i,t} = \beta x_{i,t} + \lambda_{1i}f_{1t} + \lambda_{2i}f_{2t} + \epsilon_{i,t}$ $x_{i,t} = \Lambda_{1i}f_{1t} + \Lambda_{2i}f_{2t} + \eta_{i,t}$	$\lambda_{1i} \sim N(1.0, 0.2)$ $\lambda_{2i} \sim N(0.5, 0.5)$ $\Lambda_{1i} \sim N(0.5, 0.5)$ $\Lambda_{2i} = 0$	$z_{1i} \sim N(1.0, 0.5)$ $z_{2i} \sim N(0.2, 0.5)$ $z_{3i} \sim N(0.5, 0.5)$ $z_{4i} \sim N(0.5, 0.5)$
E4	2	$y_{i,t} = \beta x_{i,t} + \lambda_{1i}f_{1t} + \lambda_{2i}f_{2t} + \epsilon_{i,t}$ $x_{i,t} = \Lambda_{1i}f_{1t} + \Lambda_{2i}f_{2t} + \eta_{i,t}$	$\lambda_{1i} = z_{1i} + 0.5\tau_{\lambda 1i}$ $\lambda_{2i} = z_{2i} + 0.5\tau_{\lambda 2i}$ $\Lambda_{1i} = z_{3i} + 0.5\tau_{\Lambda 1i}$ $\Lambda_{2i} = z_{4i} + 0.5\tau_{\Lambda 2i}$	$z_{1i} \sim N(1.0, 0.5)$ $z_{2i} \sim N(0, 0.5)$ $z_{3i} \sim N(0.5, 0.5)$ $z_{4i} \sim N(0, 0.5)$
E5	3	$y_{i,t} = \beta x_{i,t} + \lambda_{1i}f_{1t} + \lambda_{2i}f_{2t} + \lambda_{3i}f_{3t} + \epsilon_{i,t}$ $x_{i,t} = \Lambda_{1i}f_{1t} + \Lambda_{2i}f_{2t} + \Lambda_{3i}f_{3t} + \eta_{i,t}$	$\lambda_{1i} \sim N(1.0, 0.2)$ $\lambda_{2i} \sim N(0.5, 0.5)$ $\lambda_{3i} \sim N(0, 0.5)$ $\Lambda_{1i} \sim N(0.5, 0.5)$ $\Lambda_{2i} \sim N(0.2, 0.5)$ $\Lambda_{3i} = 0$	$z_{1i} \sim N(1.0, 0.5)$ $z_{2i} \sim N(0.2, 0.5)$ $z_{3i} \sim N(0.5, 0.5)$ $z_{4i} \sim N(0.5, 0.5)$
E6	2	$y_{i,t} = \beta x_{i,t} + \lambda_{1i}f_{1t} + \lambda_{2i}f_{2t} + \epsilon_{i,t}$ $x_{i,t} = \Lambda_{1i}f_{1t} + \Lambda_{2i}f_{2t} + \eta_{i,t}$	$\lambda_{1i} = z_{1i} + z_{4i} + \tau_{\lambda 1i}$ $\lambda_{2i} = z_{1i} - 5z_{3i} + \tau_{\lambda 2i}$ $\Lambda_{1i} = z_{3i} + \tau_{\Lambda 1i}$ $\Lambda_{2i} = z_{3i} - z_{4i} + \tau_{\Lambda 2i}$	$z_{1i} \sim N(1.0, 0.5)$ $z_{2i} \sim N(0.5, 0.5)$ $z_{3i} \sim N(0.2, 0.5)$ $z_{4i} \sim N(0.2, 0.5)$
E7	2	$y_{i,t} = \beta x_{i,t} + \lambda_{1i}f_{1t} + \lambda_{2i}f_{2t} + \epsilon_{i,t}$ $x_{i,t} = \Lambda_{1i}f_{1t} + \Lambda_{2i}f_{2t} + \eta_{i,t}$	$\lambda_{1i} = z_{1i} - 5z_{2i} + \tau_{\lambda 1i}$ $\lambda_{2i} = z_{1i} - 2z_{3i} + \tau_{\lambda 2i}$ $\Lambda_{1i} = z_{1i} - 2z_{3i} + \tau_{\Lambda 1i}$ $\Lambda_{2i} = z_{4i} + \tau_{\Lambda 2i}$	$z_{1i} \sim N(1.0, 0.5)$ $z_{2i} \sim N(0.2, 0.5)$ $z_{3i} \sim N(0.5, 0.5)$ $z_{4i} \sim N(0, 0.5)$

Notes: $(\tau_{\lambda 1i}, \tau_{\lambda 2i}, \tau_{\Lambda 1i}, \tau_{\Lambda 2i})' \sim N(\mathbf{1}_4, I_4)$, where $\mathbf{1}_4 = (1, 1, 1, 1)'$.

Table E1: Bias and bias correction.

		Bias $\times \sqrt{NT}$						Size								
N	T	PC	Theory	BA	CCE	Theory	BA	C ³ E	Theory	BA	PC	BA	CCE	BA	C ³ E	BA
30	30	2.370	2.201	0.342	0.642	0.713	0.016	0.249	0.311	-0.014	57.9	10.4	11.3	9.5	8.0	12.2
50	30	1.801	1.702	0.168	0.245	0.267	-0.018	0.135	0.200	-0.060	40.8	8.5	6.7	8.4	7.7	12.0
100	30	1.357	1.325	0.074	0.243	0.273	-0.023	0.104	0.103	-0.067	27.4	7.1	6.1	6.6	7.1	11.1
200	30	1.019	0.977	0.064	0.119	0.122	-0.003	0.024	0.030	-0.008	18.2	6.2	6.0	6.1	6.5	10.2
30	50	2.988	2.841	0.362	0.826	0.920	0.021	0.311	0.401	-0.023	78.5	9.7	13.9	9.2	8.3	9.9
50	50	2.282	2.197	0.165	0.316	0.344	-0.022	0.174	0.259	-0.075	59.2	7.3	6.7	7.3	6.3	8.9
100	50	1.743	1.710	0.080	0.326	0.352	-0.016	0.157	0.134	-0.054	39.6	6.3	6.6	6.1	6.3	8.5
200	50	1.274	1.261	0.036	0.136	0.158	-0.021	0.010	0.039	-0.029	24.8	5.9	6.0	6.3	6.7	8.2
30	100	4.182	4.018	0.461	1.169	1.301	0.025	0.451	0.567	-0.027	96.4	10.0	23.4	8.9	8.4	8.6
50	100	3.201	3.107	0.203	0.464	0.487	-0.010	0.264	0.366	-0.090	86.1	7.7	8.3	7.4	7.0	7.5
100	100	2.453	2.419	0.094	0.473	0.498	-0.009	0.224	0.189	-0.075	66.7	6.5	8.2	6.5	6.5	7.0
200	100	1.813	1.784	0.057	0.227	0.224	0.005	0.050	0.055	-0.008	43.5	6.1	5.6	5.6	5.7	6.3
30	200	5.890	5.682	0.625	1.644	1.841	0.032	0.629	0.802	-0.044	100.0	12.1	39.0	9.4	11.1	8.1
50	200	4.476	4.394	0.230	0.646	0.689	-0.024	0.365	0.517	-0.136	98.9	6.7	9.6	6.3	6.8	7.0
100	200	3.484	3.421	0.144	0.694	0.705	0.012	0.339	0.267	-0.089	92.1	6.5	11.5	6.2	6.9	6.0
200	200	2.539	2.522	0.051	0.314	0.316	-0.002	0.067	0.077	-0.012	69.9	5.7	6.4	5.6	5.6	5.9
93	30	1.456	1.373	0.129	0.277	0.278	0.008	0.136	0.086	-0.087	30.8	7.5	7.1	7.4	7.4	11.7
184	50	1.345	1.322	0.049	0.202	0.210	-0.004	0.055	0.043	-0.033	26.8	6.0	6.0	6.0	5.8	7.7
464	100	1.211	1.206	0.016	0.108	0.121	-0.012	-0.015	-0.012	-0.017	22.8	5.7	5.7	5.8	5.6	6.7

Notes: "PC" and "CCE" refer to the principal components estimator of Bai (2009), and the common correlated effects estimator of Pesaran (2006), respectively. "C³E" is the modified CCE estimator proposed in this paper. "Theory" and "BA" refer to the theoretical prediction and bias-adjusted version of each estimator. The theoretical bias of the CCE and C³E estimators are obtained from Theorem 1. In case of the PC estimator the theoretical prediction is taken from Bai (2009).

Table E2: All conditions of PC, CCE and C³E are satisfied.

		Bias $\times 100$								Size							
N	T	LS	PC	CCE	C ³ E	C ³ E 1	C ³ ES	C ³ ES 1	LS	PC	CCE	C ³ E	C ³ E 1	C ³ ES	C ³ ES 1		
30	30	44.614	7.758	-0.732	0.275	0.290	0.289	0.202	100.0	57.7	6.8	12.7	15.6	12.8	11.7		
50	30	41.346	4.890	0.089	0.178	0.142	0.179	0.148	100.0	44.2	6.4	10.7	12.6	10.7	10.4		
100	30	42.770	2.391	0.107	0.166	0.163	0.166	0.153	100.0	26.5	5.7	10.5	12.0	10.5	9.8		
200	30	39.145	1.068	-0.073	0.043	0.044	0.043	0.043	100.0	14.1	6.6	10.1	11.5	10.1	9.7		
30	50	44.879	7.620	-0.713	0.269	0.308	0.539	-0.534	100.0	76.6	7.2	10.1	12.0	9.2	7.3		
50	50	41.668	4.782	0.072	0.232	0.208	0.421	0.107	100.0	61.6	5.8	8.7	9.8	8.3	7.0		
100	50	42.529	2.328	0.067	0.142	0.131	0.142	0.126	100.0	36.1	5.5	7.9	8.1	7.9	7.6		
200	50	39.515	1.048	-0.093	0.042	0.039	0.042	0.024	100.0	18.5	5.4	7.3	8.2	7.3	7.2		
30	100	44.896	7.573	-0.685	0.315	0.359	0.598	-0.697	100.0	96.2	7.0	8.5	9.5	10.1	7.1		
50	100	41.733	4.775	0.108	0.254	0.220	0.621	0.108	100.0	88.6	5.6	7.6	7.9	11.0	5.6		
100	100	42.646	2.321	0.081	0.155	0.150	0.274	0.077	100.0	61.9	5.7	7.0	7.8	7.3	5.7		
200	100	39.517	1.058	-0.072	0.049	0.049	0.049	0.032	100.0	31.8	5.3	6.1	6.5	6.1	6.4		
30	200	44.974	7.544	-0.676	0.319	0.361	0.586	-0.678	100.0	99.9	8.9	8.3	8.8	13.1	8.9		
50	200	41.721	4.724	0.069	0.228	0.203	0.580	0.069	100.0	99.1	5.8	7.4	7.8	16.4	5.8		
100	200	42.703	2.296	0.074	0.145	0.144	0.239	0.074	100.0	87.9	5.6	6.5	6.9	8.5	5.6		
200	200	39.535	1.045	-0.081	0.041	0.044	0.051	-0.081	100.0	53.7	5.2	5.6	5.7	6.3	5.2		
93	30	43.193	2.565	0.056	0.176	0.182	0.176	0.143	100.0	27.0	6.5	10.5	12.1	10.5	10.4		
184	50	39.665	1.157	-0.074	0.044	0.040	0.044	0.020	100.0	20.5	6.0	7.8	8.5	7.8	7.9		
464	100	42.351	0.446	-0.115	0.002	0.003	0.002	-0.007	100.0	16.1	5.4	5.6	5.9	5.6	5.7		

Notes: "LS" and "C³ES" refers to the least squares and C³E estimator based on IC selection of the combinations. See Table E1 for an explanation of the rest.

Table E3: The combinations are uncorrelated with the loadings.

N	T	Bias $\times 100$							Size						
		LS	PC	CCE	C ³ E	C ³ E 1	C ³ ES	C ³ ES 1	LS	PC	CCE	C ³ E	C ³ E 1	C ³ ES	C ³ ES 1
30	30	25.648	10.510	-0.972	-0.520	-0.144	-0.490	-0.290	99.9	79.1	6.7	12.6	15.5	12.4	11.8
50	30	29.162	6.250	0.391	0.000	0.291	0.002	0.323	100.0	61.1	6.3	10.4	12.2	10.4	10.8
100	30	30.077	3.349	0.175	0.121	0.172	0.121	0.184	100.0	43.2	6.1	9.7	11.8	9.7	9.7
200	30	31.317	1.756	0.042	0.056	0.037	0.056	0.046	100.0	27.3	6.4	9.8	11.5	9.8	9.5
30	50	25.743	10.057	-0.965	-0.532	-0.162	-0.350	-0.753	100.0	92.6	7.9	10.3	11.5	8.7	8.1
50	50	29.482	6.023	0.363	0.008	0.323	0.118	0.355	100.0	79.7	6.5	8.5	9.7	8.0	7.0
100	50	30.125	3.198	0.156	0.109	0.158	0.109	0.168	100.0	58.5	5.6	7.8	8.7	7.8	7.6
200	50	31.565	1.666	0.020	0.056	0.039	0.056	0.028	100.0	38.6	5.5	7.5	8.1	7.5	7.6
30	100	25.899	9.843	-0.941	-0.508	-0.128	-0.168	-0.906	100.0	99.8	8.5	8.9	9.1	8.1	8.4
50	100	29.613	5.934	0.395	0.057	0.359	0.541	0.394	100.0	97.4	6.4	7.1	8.5	12.0	6.4
100	100	30.284	3.131	0.162	0.108	0.159	0.144	0.161	100.0	85.3	6.0	7.0	8.0	6.9	6.0
200	100	31.714	1.639	0.039	0.060	0.045	0.062	0.045	100.0	62.6	5.3	6.3	6.9	6.3	6.3
30	200	25.946	9.725	-0.924	-0.482	-0.117	-0.092	-0.909	100.0	100.0	11.0	8.8	7.8	9.3	11.0
50	200	29.623	5.840	0.364	0.014	0.329	0.543	0.364	100.0	100.0	7.0	6.9	8.6	17.3	7.0
100	200	30.367	3.085	0.157	0.107	0.157	0.199	0.157	100.0	98.4	5.9	6.3	7.0	7.4	5.9
200	200	31.690	1.611	0.032	0.057	0.041	0.070	0.032	100.0	88.2	5.0	5.9	6.0	5.7	5.0
93	30	30.090	3.619	0.139	0.053	0.136	0.053	0.141	100.0	45.7	6.1	10.0	12.3	10.0	10.3
184	50	31.520	1.819	0.042	0.073	0.061	0.073	0.054	100.0	41.0	5.9	7.9	8.3	7.9	7.7
464	100	30.759	0.700	0.006	0.010	0.009	0.010	0.009	100.0	32.3	4.8	5.9	5.9	5.9	5.5

Notes: "LS" and "C³ES" refers to the least squares and C³E estimator based on IC selection of the combinations. See Table E1 for an explanation of the rest.

Table E4: Condition (5) is not satisfied but loadings are iid.

		Bias $\times 100$								Size							
N	T	LS	PC	CCE	C ³ E	C ³ E 1	C ³ ES	C ³ ES 1	LS	PC	CCE	C ³ E	C ³ E 1	C ³ ES	C ³ ES 1		
30	30	15.028	5.993	-3.013	0.246	0.236	0.250	0.119	96.8	40.1	14.0	12.4	15.2	12.3	12.2		
50	30	15.056	3.531	-2.810	0.104	0.066	0.104	0.087	99.1	27.1	18.9	10.4	12.9	10.4	10.9		
100	30	20.478	1.846	1.361	0.106	0.101	0.106	0.115	100.0	18.0	12.2	10.0	11.8	10.0	9.5		
200	30	15.474	0.876	-1.105	0.018	0.016	0.018	0.016	100.0	11.6	15.5	10.3	11.8	10.3	9.8		
30	50	15.019	5.866	-3.027	0.180	0.166	0.152	-1.749	99.8	56.9	19.5	10.4	11.7	8.9	14.4		
50	50	15.194	3.404	-2.886	0.141	0.114	0.106	-0.911	100.0	38.1	30.3	8.4	9.3	7.8	15.1		
100	50	20.546	1.768	1.318	0.075	0.070	0.075	0.086	100.0	23.8	15.2	7.7	8.1	7.7	7.4		
200	50	15.490	0.838	-1.150	0.019	0.017	0.019	0.008	100.0	13.2	22.3	7.1	7.9	7.1	7.3		
30	100	15.037	5.776	-3.089	0.233	0.229	0.201	-2.222	100.0	83.8	38.9	8.2	9.0	6.7	27.1		
50	100	15.278	3.384	-2.849	0.166	0.134	0.167	-2.625	100.0	63.3	52.1	7.6	7.7	6.5	46.7		
100	100	20.621	1.744	1.319	0.084	0.080	0.117	1.103	100.0	39.7	25.5	7.1	7.4	6.2	21.0		
200	100	15.594	0.845	-1.130	0.027	0.025	0.025	-0.005	100.0	22.3	37.4	6.2	6.6	6.1	6.9		
30	200	15.070	5.746	-3.063	0.247	0.236	0.205	-2.235	100.0	98.8	68.1	7.8	8.5	7.0	45.2		
50	200	15.264	3.343	-2.876	0.141	0.120	0.145	-2.655	100.0	89.1	82.4	7.5	7.9	6.7	75.0		
100	200	20.670	1.726	1.342	0.074	0.073	0.121	1.225	100.0	66.6	47.7	6.3	6.4	6.0	41.3		
200	200	15.539	0.828	-1.137	0.021	0.023	0.014	-1.097	100.0	38.0	63.4	5.2	5.8	5.0	60.1		
93	30	20.345	2.018	0.985	0.136	0.132	0.136	0.137	100.0	19.6	9.3	10.6	12.1	10.6	10.7		
184	50	14.603	0.908	-1.348	0.018	0.016	0.018	0.004	100.0	14.8	26.8	7.9	8.6	7.9	8.0		
464	100	14.054	0.359	-1.947	-0.001	-0.001	-0.001	-0.005	100.0	12.0	98.4	5.7	6.0	5.7	5.6		

Notes: "LS" and "C³ES" refers to the least squares and C³E estimator based on IC selection of the combinations. See Table E1 for an explanation of the rest.

Table E5: $m + 1 < r$ but loadings are iid and independent of the combinations.

		Bias $\times 100$								Size							
N	T	LS	PC	CCE	C ³ E	C ³ E 1	C ³ ES	C ³ ES 1	LS	PC	CCE	C ³ E	C ³ E 1	C ³ ES	C ³ ES 1		
30	30	42.215	14.162	1.877	1.722	-0.121	1.735	0.118	100.0	94.0	7.1	15.3	14.6	15.3	12.0		
50	30	44.695	8.816	0.590	0.779	-0.154	0.779	-0.105	100.0	84.2	5.3	11.9	12.9	11.9	10.3		
100	30	42.529	4.675	0.504	0.362	0.177	0.362	0.298	100.0	66.2	6.3	10.8	12.4	10.8	10.4		
200	30	42.124	2.352	0.642	0.345	0.286	0.345	0.367	100.0	43.3	7.8	11.0	12.5	11.0	10.8		
30	50	42.353	13.423	1.894	1.670	-0.155	1.779	0.284	100.0	99.4	8.0	14.2	11.3	14.5	8.4		
50	50	44.963	8.380	0.633	0.817	-0.134	0.822	-0.031	100.0	96.3	5.6	10.0	9.1	9.9	7.7		
100	50	42.819	4.415	0.485	0.381	0.194	0.381	0.301	100.0	84.1	6.4	8.1	8.7	8.1	7.9		
200	50	42.216	2.257	0.668	0.378	0.319	0.378	0.397	100.0	58.7	9.1	9.0	9.6	9.0	9.4		
30	100	42.556	13.082	1.909	1.753	-0.070	2.020	0.663	100.0	100.0	13.6	18.0	9.4	21.3	9.0		
50	100	45.194	8.084	0.628	0.804	-0.142	0.822	0.115	100.0	99.9	6.4	10.8	8.1	10.4	7.2		
100	100	42.951	4.250	0.489	0.391	0.200	0.393	0.346	100.0	97.9	7.1	8.6	7.7	8.6	8.0		
200	100	42.623	2.161	0.646	0.363	0.298	0.363	0.372	100.0	83.4	12.8	9.8	9.0	9.8	9.5		
30	200	42.647	12.878	1.906	1.739	-0.106	2.116	0.826	100.0	100.0	22.9	27.0	8.6	35.2	12.9		
50	200	45.272	7.927	0.635	0.785	-0.166	0.819	0.240	100.0	100.0	7.9	12.4	7.6	12.8	7.5		
100	200	42.966	4.156	0.489	0.382	0.190	0.390	0.419	100.0	100.0	9.4	9.5	7.2	9.7	10.8		
200	200	42.705	2.122	0.670	0.362	0.300	0.363	0.428	100.0	98.7	20.7	11.5	9.2	11.5	12.9		
93	30	42.519	5.048	0.611	0.642	0.406	0.642	0.478	100.0	69.0	6.4	11.0	12.4	11.0	10.8		
184	50	43.774	2.457	0.652	0.423	0.340	0.423	0.431	100.0	62.5	8.9	9.1	9.4	9.1	9.3		
464	100	41.449	0.913	0.014	0.053	0.026	0.053	0.025	100.0	49.1	5.3	6.3	6.6	6.3	5.9		

Notes: "LS" and "C³ES" refers to the least squares and C³E estimator based on IC selection of the combinations. See Table E1 for an explanation of the rest.

Table E6: Condition (5) is not satisfied and loadings are non-iid.

N	T	Bias $\times 100$								Size							
		LS	PC	CCE	C ³ E	C ³ E 1	C ³ ES	C ³ ES 1	LS	PC	CCE	C ³ E	C ³ E 1	C ³ ES	C ³ ES 1		
30	30	-46.524	7.297	-35.300	0.121	0.096	0.117	-3.205	99.9	53.0	100.0	12.3	15.3	12.2	27.7		
50	30	-29.765	4.342	-28.934	0.101	0.058	0.102	-1.200	95.4	37.3	99.9	10.4	11.9	10.4	16.0		
100	30	-27.598	2.265	-5.207	0.067	0.056	0.067	-0.490	95.3	23.8	68.9	9.9	11.3	9.9	12.7		
200	30	-31.547	1.228	-14.184	0.020	0.019	0.020	-0.421	97.2	16.7	100.0	10.5	11.3	10.5	12.9		
30	50	-47.002	7.221	-35.369	0.073	0.073	0.103	-9.996	100.0	73.5	100.0	10.0	11.5	9.2	50.9		
50	50	-29.932	4.277	-29.026	0.137	0.137	0.172	-3.310	99.2	53.9	100.0	8.6	10.1	8.1	25.2		
100	50	-27.323	2.209	-5.295	0.038	0.032	0.038	-0.561	98.8	33.5	87.4	7.7	8.2	7.7	12.0		
200	50	-32.066	1.201	-14.296	0.028	0.021	0.028	-0.443	99.5	22.5	100.0	7.0	7.7	7.0	13.2		
30	100	-47.083	7.241	-35.493	0.126	0.131	0.126	-17.769	100.0	95.3	100.0	8.2	9.2	7.5	85.5		
50	100	-29.951	4.304	-29.179	0.161	0.154	0.297	-13.797	100.0	82.7	100.0	7.3	7.8	7.1	68.8		
100	100	-27.404	2.222	-5.274	0.049	0.044	0.096	-3.006	100.0	58.3	99.0	7.0	7.4	6.4	58.1		
200	100	-31.931	1.212	-14.253	0.033	0.031	0.033	-0.493	100.0	39.5	100.0	6.4	6.7	6.4	17.7		
30	200	-47.234	7.205	-35.596	0.120	0.117	0.112	-20.343	100.0	100.0	100.0	7.1	8.1	7.2	99.1		
50	200	-29.975	4.272	-29.281	0.135	0.141	0.278	-14.599	100.0	97.6	100.0	7.5	7.9	8.3	85.9		
100	200	-27.452	2.203	-5.280	0.040	0.038	0.127	-3.233	100.0	86.0	100.0	6.1	6.4	6.3	70.3		
200	200	-32.165	1.201	-14.278	0.030	0.031	0.066	-6.955	100.0	65.1	100.0	5.7	5.9	5.6	75.7		
93	30	-28.469	2.409	-7.179	0.097	0.101	0.097	-0.622	96.6	24.6	84.3	10.5	12.2	10.5	13.7		
184	50	-34.199	1.312	-13.404	0.023	0.019	0.023	-0.415	99.7	24.9	100.0	7.0	8.3	7.0	12.5		
464	100	-45.955	0.500	-65.327	0.005	0.004	0.005	-0.187	100.0	18.8	100.0	5.6	5.8	5.6	10.2		

Notes: "LS" and "C³ES" refers to the least squares and C³E estimator based on IC selection of the combinations. See Table E1 for an explanation of the rest.

Table E7: Loadings are non-iid and mean zero.

N	T	Bias $\times 100$						Size					
		LS	PC	CCE	C ³ E	C ³ E 1	C ³ ES 1	LS	PC	CCE	C ³ E	C ³ E 1	C ³ ES 1
30	30	7.368	8.019	-12.079	0.518	0.488	0.539	29.7	60.1	72.6	12.5	15.3	12.5
50	30	3.963	4.854	11.649	0.183	0.109	0.183	33.0	44.0	70.2	10.8	12.6	10.8
100	30	21.593	2.582	-3.260	0.148	0.143	0.148	98.3	29.4	28.4	10.3	11.6	10.3
200	30	19.540	1.214	-16.140	0.029	0.022	0.029	95.3	17.0	100.0	10.1	11.2	10.1
30	50	7.419	7.896	-12.165	0.464	0.437	0.589	39.2	80.1	89.1	9.8	11.7	9.1
50	50	4.068	4.789	11.887	0.220	0.173	0.246	37.9	62.4	84.3	8.9	9.5	8.9
100	50	21.284	2.526	-3.461	0.110	0.104	0.110	99.8	41.7	40.9	7.9	8.4	7.9
200	50	19.814	1.195	-16.371	0.023	0.022	0.023	99.0	22.3	100.0	7.2	7.7	7.2
30	100	7.359	7.877	-12.263	0.500	0.486	0.646	57.9	97.4	99.5	9.2	10.2	9.6
50	100	3.987	4.781	11.861	0.227	0.169	0.373	44.5	89.1	96.7	7.6	8.1	7.6
100	100	21.312	2.536	-3.427	0.117	0.112	0.188	100.0	69.4	62.1	6.9	7.7	7.0
200	100	19.740	1.200	-16.482	0.029	0.028	0.029	100.0	39.5	100.0	6.4	6.7	6.3
30	200	7.298	7.874	-12.333	0.506	0.484	0.652	81.0	100.0	100.0	9.1	9.5	10.2
50	200	3.953	4.762	11.926	0.206	0.159	0.363	57.6	99.1	99.7	7.6	7.9	8.8
100	200	21.324	2.517	-3.516	0.112	0.109	0.179	100.0	93.3	87.5	6.3	6.5	6.8
200	200	19.689	1.194	-16.558	0.022	0.024	0.032	100.0	65.5	100.0	5.0	5.4	5.0
93	30	20.139	2.680	-6.143	0.159	0.152	0.159	96.8	29.2	56.7	10.1	11.9	10.3
184	50	21.602	1.351	-13.821	0.018	0.016	0.018	99.0	25.6	100.0	7.5	8.1	7.5
464	100	15.239	0.505	-4.653	-0.001	-0.001	-0.001	99.8	19.5	97.6	5.5	5.8	5.5

Notes: "LS" and "C³ES" refers to the least squares and C³E estimator based on IC selection of the combinations. See Table E1 for an explanation of the rest.